

The international journal of science / 5 January 2023

nature

NET LOSS

Papers and patents are becoming less disruptive of the scientific status quo

Social upheaval
Will the turmoil engulfing Twitter also disrupt science?

Evolutionary odyssey
The forgotten legacy of Alfred Russel Wallace's voyage to the Amazon

Farming life
Eleven ways to help croplands reduce nitrogen pollution

VISIT…

LANZAROTE Caliente.com

Nature.2023.01.07

[Sat, 07 Jan 2023]

- **This Week**
- **News in Focus**
- **Opinion**
- **Work**
- **Research**
- **Amendments & Corrections**

This Week

- **There’s no room for COVID complacency in 2023** [23 Dec 2022]
  • Stark scenes from China show the pandemic is far from over. One solution is a laser-like focus on strengthening public-health systems.
- **Is my study useless? Why researchers need methodological review boards** [03 Jan 2023]
  • Making researchers account for their methods before data collection is a long-overdue step.

- EDITORIAL
- 23 December 2022

There’s no room for COVID complacency in 2023

Stark scenes from China show the pandemic is far from over. One solution is a laser-like focus on strengthening public-health systems.



People queue outside a Beijing pharmacy for COVID-19 treatments. Models suggest China could see up to one million deaths in 2023.Credit: Kevin Frayer/Getty

In many places, life took on a semblance of pre-COVID normality in 2022, as countries shed pandemic-control measures. Governments ended lockdowns,

reopened schools and scaled back or abandoned mask-wearing mandates. International travel resumed.

There were optimistic proclamations, too. In January, Danish Prime Minister Mette Frederiksen declared that SARS-CoV-2 no longer poses a threat to society. In September, US President Joe Biden remarked during an interview that the pandemic was over. Even Tedros Adhanom Ghebreyesus, director-general of the World Health Organization (WHO), has expressed hope that COVID-19's designation as a global emergency will end in 2023.

This belies the devastation that the disease continues to cause. The starkest example is in China, one of the last countries to ease pandemic-control measures in the face of the fast-spreading Omicron variant. Scenes emerging from Chinese hospitals now are reminiscent of the havoc that Omicron wrought in Hong Kong nearly a year ago. China might have seen widespread transmission regardless of whether President Xi Jinping had dropped the zero-COVID policy in December. But models suggest that the country faces the prospect of up to one million deaths over the next year, not to mention widespread workplace absences and disruptions to the Chinese — and global — economy.

Most people in China are immunologically unprepared for Omicron, the dominant strain now in circulation. They have had no exposure to any SARS-CoV-2 variant and, if vaccinated, have received vaccines only against the virus's original strain. China is likely to discover what other countries with limited previous exposure to the virus have found over the past year: that there won't be a single 'exit' wave to mark the lifting of pandemic restrictions. Further waves of infection and death are likely to follow, either from new variants that arise in the population, or from variants imported as the country opens its borders to visitors.

Renewed responses needed

Elsewhere, repeated surges in infection and death are giving way to a constant thrum of loss, as well as debilitation caused by long COVID. A focus on COVID-19 has also affected the fights against AIDS, malaria and tuberculosis. Although precise counts are difficult to obtain, overall death

rates in many countries remain higher than before COVID-19 hit.

COVID-19 vaccination rates have stalled in many nations. In some, the uptake of boosters has been dismal, even though these substantially reduce death and severe illness.

One path to renewing vaccination efforts lies with technology. Development of mucosal vaccines is under way. These are designed to be delivered through the nose or mouth and it’s hoped they can trigger sterilizing immunity that blocks transmission — not just severe illness. China has approved an inhalable booster dose and a nasal vaccine, and India a two-dose nasal-drop primary vaccine. Iran and Russia have also each approved a mucosal vaccine. But researchers are awaiting data to check whether any of these deliver on their promise of stopping SARS-CoV-2.

One thing that could shake COVID complacency is the emergence of one or more ‘variants of concern’ (VoCs). New variants of the virus will emerge over the next year, as they did in 2022. But a VoC designation (and a corresponding Greek letter from the WHO) will be given only if a variant is better at evading the immune system, causes more severe disease or is much more transmissible than those currently circulating. A new VoC must spur action to ensure that fully vaccinated people — especially those who are older or immunocompromised — receive booster doses.

A new variant must also prompt redoubled vaccination efforts in lower-income nations. Global collaborations, such as COVAX, were established to deliver vaccines equitably. But they faltered as wealthy nations prioritized vaccinating their own populations. Too often, vaccines for low- and middle-income countries (LMICs) were delivered sporadically and close to their date of expiry, exacerbating the challenge of rolling them out in places with limited health-care infrastructure.

The result is that only one-quarter of people in low-income countries have received at least one dose of a coronavirus vaccine. Many low-income countries need to get back to tackling neglected priorities such as malaria, tuberculosis and infant mortality, all of which were sidelined as the worst of the pandemic swept through. But ignoring COVID-19’s continued toll risks stymieing these efforts, too.

The global community must reckon with the politics and power dynamics that undermined initiatives to ensure that all nations had access to vaccines when they needed them. Unless that happens, future global agreements could be similarly undermined in times of crisis. In May, the WHO's intergovernmental negotiating body will deliver a progress report on deliberations over an international instrument — the nearest thing to a treaty — on pandemic preparedness and response. Countries that missed out on timely access to COVID-19 vaccines, tests and treatments will be arguing that the agreement should ensure more equitable access to resources when the next pandemic threat emerges.

But as attention moves to preparations for 'disease X' — the as-yet-unknown pathogen that could cause the next pandemic — COVID complacency is inflicting death by a thousand cuts on health-care systems reeling from the past three years. The public-health community must continue to strengthen vaccine-manufacturing capacity in LMICs. And it mustn't forget what experience has shown since 2020: that health-care systems under stress are little able to deal with new threats.

Nature **613**, 7 (2023)

doi: https://doi.org/10.1038/d41586-022-04476-9

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-022-04476-9

- WORLD VIEW
- 03 January 2023

Is my study useless? Why researchers need methodological review boards


Making researchers account for their methods before data collection is a long-overdue step.

- Daniël Lakens

Should researchers have the freedom to perform research that is a waste of time? Currently, the answer is a resounding ‘yes’. Or at least, no one stops to ask whether there are obvious methodological and statistical flaws in a proposed study that will make it useless from the get-go: a sample size that’s simply too small to test a hypothesis, for example.

In my role as chair of the central ethical review board at Eindhoven

University of Technology in the Netherlands, I’ve lost count of the number of times that a board member has remarked that, although we’re not supposed to comment on non-ethical issues, the way a study has been designed means it won’t yield any informative data. And yet we routinely wait until peer review — after the study has been done — to identify flaws that can’t then be corrected.

In my own department at Eindhoven, we’ve been trialling a different approach. Five years ago, we instituted a local review board that also evaluates proposed methods. Although some colleagues found this extra hurdle frustrating at first, the improvements in study quality have led them to accept it. It’s time to make dedicated methodological review boards a standard feature at universities and other research institutions, as institutional review boards are.



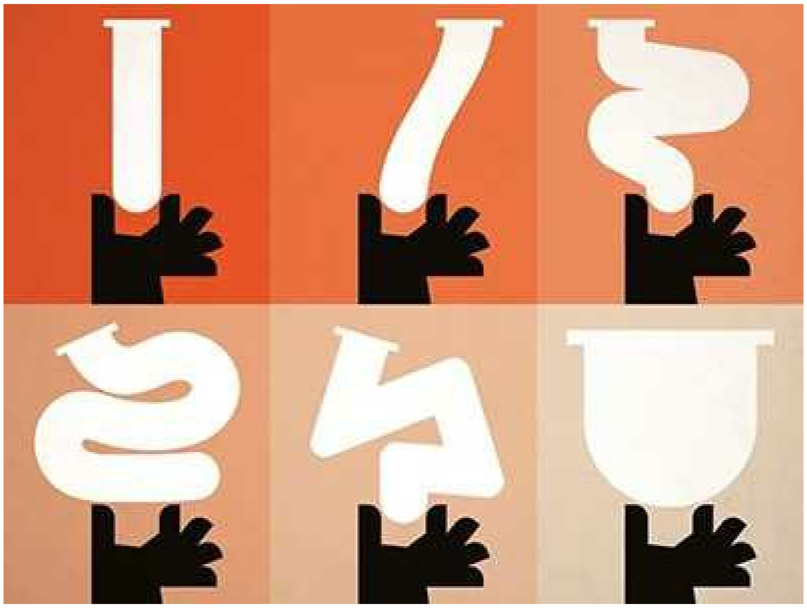
To fix peer review, break it into stages

Some medical trials, animal studies, grant applications and institutes around the world already review methods. For example, in stage one peer review when submitting a registered report, or peer review of a clinical trial protocol, reviewers comment on the study design before data collection. If a study has already passed such hurdles, a methodological review board will no longer need to assess it. That said, there are signs the existing system needs tightening up. The journal *Trials* has used clinical-trial protocol reviewers

since September 2019, for example: it has found some items of methodological information missing in up to 56% of protocols (R. Qureishi *et al. Trials* **23**, 359; 2022). Normal peer review had not flagged this.

To be clear, I do not propose that reviewers debate matters such as frequentist versus Bayesian philosophies of statistics. Instead, the focus should be on basic design flaws that cannot be corrected after data collection, with the goal of ensuring that the data can inform the statistical hypothesis being tested. For one thing, reviewers could check that researchers will collect sufficient data and be able to make the causal inferences they desire (by ensuring that the sample is representative of the target population, for example). Reviewers could deter researchers from performing too many exploratory analyses while reporting only some of them, and help to plan studies that will yield informative results even if the hypothesized effect is absent. They should also check that researchers follow disciplinary reporting guidelines where available.

Critics might worry that methodological reviewers will abuse their power and prohibit certain contentious methods or manipulations — for example, the implicit association test. (Many critics say the test does not measure implicit associations at all.) But the methodological review I propose is not about whether measures and manipulations are valid. Discussions about possible confounding variables and bad measures are too complex and, in my view, are best resolved in the literature.



[Five keys to writing a reproducible lab protocol]

The most contentious issue in methodological review is that boards have the power, in principle, to bar a proposed study from proceeding. Since we introduced methodological reviews as part of the ethics review process in our department, this has never happened. The methods can usually be adjusted to fit the question, or the question can be rephrased to fit the methods. Over time, we have asked colleagues to increase the level of detail about their study design and analysis methods. This led to more clearly specified statistical hypotheses. People have increasingly used sequential analyses — an efficient way to collect data, involving adding interim analyses at predetermined sample sizes to determine if the data are now sufficient — because writing a sample-size justification makes them realize they are uncertain about the sample size they need.

There are other advantages, too. Introducing methodological review boards might create dedicated career paths for specialists in research methods — individuals who add great value to the research enterprise, but whose contributions rarely get them tenure in the current system.

As part of roll-out, studies should test whether or not methodological review improves the quality of research. Proposals could be randomly assigned either to receive methodological review or to a control condition (for example, one in which the review goes ahead but the results are kept private).

If the review does not sufficiently improve studies at a research institution, it should be abandoned.

It takes time for researchers to learn how to provide all the required information, so methodological review boards should be implemented slowly. Reviewers might initially provide suggestions, without requiring their incorporation in the final study design. Alternatively, the first stage could focus on studies that collect data from vulnerable populations or that require large resources or might have immediate societal impact. And reviewers should actively help researchers to improve their study proposals.

The goal is not to gatekeep, but to improve. As researchers learn better practices, all will benefit from the more informative studies they design.

Nature **613**, 9 (2023)

doi: https://doi.org/10.1038/d41586-022-04504-8

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-022-04504-8

News in Focus

- ## The science events to watch for in 2023 [19 Dec 2022]
  • Moon landings, mRNA vaccines and climate finance are among the developments set to shape research in the coming year.
- ## After AlphaFold: protein-folding contest seeks next big breakthrough [13 Dec 2022]
  • Two years after DeepMind's revolutionary AI swept a competition for predicting protein structures, researchers are building on AlphaFold's success.
- ## Are we in the Anthropocene? Geologists could define new epoch for Earth [13 Dec 2022]
  • Researchers have zeroed in on nine sites that could describe a new geological time, marked by pollution and other signs of human activity.
- ## The 'breakthrough' obesity drugs that have stunned researchers [04 Jan 2023]
  • A class of drugs that quash hunger have shown striking results in trials and in practice. But can they help all people with obesity — and conquer weight stigma?
- ## Twitter changed science — what happens now it's in turmoil? [20 Dec 2022]
  • The microblogging platform has transformed research communication, but its future is in doubt.

- NEWS
- 19 December 2022

The science events to watch for in 2023

Moon landings, mRNA vaccines and climate finance are among the developments set to shape research in the coming year.

- Miryam Naddaf



A technician installs fibre-optic cables at the Vera C. Rubin Observatory,

which is due to take its first images in 2023. Credit: H Stockebrand/Rubin Obs/NSF/AURA

Next-generation vaccines

Following the successful deployment of mRNA vaccines during the COVID-19 pandemic, a host of them are in development. BioNTech in Mainz, Germany, is expected to initiate first-in-human trials for mRNA vaccines against malaria, tuberculosis and genital herpes in the coming weeks. BioNTech is also collaborating with Pfizer, based in New York City, to trial an mRNA-based vaccine candidate to reduce the rate of shingles. Moderna in Cambridge, Massachusetts, also has mRNA vaccine candidates for the viruses that cause genital herpes and shingles.



Nature’s biggest news stories of 2022

In November, BioNTech and Pfizer began their phase I trial of an mRNA vaccine designed to protect against both COVID-19 and influenza. The combination vaccine contains mRNA strands encoding binding proteins for SARS-CoV-2, Omicron BA.4/BA.5 and four influenza variants.

Others are investigating the possibility of delivering COVID-19 vaccines using fast-acting nasal sprays. These sprays have been effective in animals, but the road to human trials could be long.

Advanced stargazing

The first images from the James Webb Space Telescope (JWST) left the world in awe. Some of the JWST's findings on the early universe were published this year, and astronomers will continue to share the telescope's results and discoveries on the evolution of galaxies in the year ahead.

The Euclid space telescope, under development by the European Space Agency (ESA), is intended to orbit the Sun for six years and capture photos to create a 3D map of the Universe; it is due to blast off in 2023. So is the Japan Aerospace Exploration Agency's X-ray Imaging and Spectroscopy Mission — an Earth-orbiting satellite that will detect X-ray radiation from distant stars and galaxies.

The Vera C. Rubin Observatory in Chile is due to start taking images in July 2023. The telescope, which has a special three-mirror design and a camera containing more than three billion pixels of solid-state detectors, will be able to scan the entire southern sky in just three nights. And the world's largest steerable telescope — the Xinjiang Qitai Radio Telescope (QTT) in Xinjiang, China — will be switched on. The QTT's fully steerable dish, spanning 110 metres, will enable it to observe 75% of the stars in the sky at any given time.



The Jupiter Icy Moons Explorer mission (illustration) will study the planet and its moons Ganymede, Europa and Callisto.Credit: spacecraft: ESA/ATG medialab; Jupiter: NASA/ESA/J. Nichols (University of Leicester); Ganymede: NASA/JPL; Io: NASA/JPL/University of Arizona; Callisto and Europa: NASA/JPL/DLR

Pathogen watchlist

The World Health Organization is expected to publish a revised list of priority pathogens. Around 300 scientists will review the evidence on more than 25 viral and bacterial families to identify pathogens that could potentially cause future outbreaks. Research and development road maps for each priority pathogen will outline knowledge gaps, set research priorities and guide the development of vaccines, treatments and diagnostic tests.

Moon missions

Just as NASA's uncrewed Orion capsule splashed back down to Earth on 11 December, three other missions were launched Moonwards: the United Arab Emirates' Rashid rover, NASA's Lunar Flashlight and the Japanese HAKUTO-R Mission 1, which will attempt a soft lunar landing in April. The Indian Space Research Organisation's third Moon-exploring mission, Chandrayaan-3, will land near the south pole in mid-2023. Next year will also see the first civilian trip to the Moon, with 11 people embarking on a 6-day private space flight aboard the SpaceX rocket Starship.

And, in April, ESA will launch its Jupiter Icy Moons Explorer (JUICE) mission, which will aim to study the environment on the giant gas planet and three of its moons.

CRISPR therapy

Next year might herald the first approval of a CRISPR gene-editing therapy, following promising results from clinical trials that used the CRISPR–Cas9 system to treat β-thalassaemia and sickle-cell disease, two genetic blood disorders. The exagamglogene autotemcel (exa-cel) treatment is being developed by the Massachusetts companies Vertex Pharmaceuticals in

Boston and CRISPR Therapeutics in Cambridge. It works by collecting a person’s own stem cells and using the CRISPR–Cas9 technology to edit the faulty gene, before infusing the cells back into the person. Vertex is expected to apply to the US Food and Drug Administration in March for approval to make exa-cel available to people with β-thalassaemia or sickle-cell disease.



The European Spallation Source will welcome its first researchers next year.Credit: Perry Nordeng/ESS

Loss and damage

The agreement on a loss-and-damage fund during the 27th United Nations Climate Change Conference (COP27) in Sharm El-Sheikh, Egypt, last month, marked an important step towards climate justice. Under the agreement, wealthy countries that have historically been responsible for high emissions will financially compensate poorer nations, which have borne the brunt of climate change. But the details still need to be worked out. A ‘transitional committee’ is expected to meet before the end of March to make

recommendations on how to arrange these funds, which will be presented to delegates from around the world during the United Nations COP28 conference in Dubai next November.

Beyond the standard model

Physicists unveiled the first results of the Muon *g* — 2 experiment in April this year, and are expected to publish more precise results in 2023. The experiment studies how short-lived particles known as muons behave in magnetic fields, and creates a sensitive test of the standard model of particle physics.

The Jiangmen Underground Neutrino Observatory in southern China will also begin searching for physics beyond the standard model, using a detector kept 700 metres underground to precisely measure the oscillation of neutrinos — electrically neutral subatomic particles.

Another highly anticipated event for particle physicists is the opening of the European Spallation Source (ESS) near Lund, Sweden. The pan-European project will generate intense neutron beams to study the structure of materials, using the most powerful linear proton accelerator ever built. The ESS will welcome its first researchers next year.



A tunnel at the underground storage facility for spent nuclear fuel in Olkiluoto, Finland.Credit: Antti Yrjonen/ZUMA Wire

Alzheimer’s drugs

In early January, US regulators will announce whether a drug that slowed down the rate of cognitive decline in a robust clinical trial[1] can become available to people with Alzheimer’s disease. Developed by the pharmaceutical company Eisai and the biotechnology firm Biogen, lecanemab is a monoclonal antibody that clears amyloid-β protein that builds up in the brain. The clinical trial included 1,795 people with early-stage Alzheimer’s and showed that lecanemab slowed mental decline by 27% compared with a placebo. However, some scientists think this is only a modest benefit and others are concerned about the safety of the drug.

Another Alzheimer’s drug, called blarcamesine — developed by Anavex Life Sciences in New York City — will continue to make its way through clinical trials. Blarcamesine activates a protein that improves the stability of neurons

and their ability to connect with each other.

Spent-fuel storage

The world's first nuclear-waste storage facility is due to begin operating next year in Olkiluoto, an island off the southwest coast of Finland. The Finnish government approved construction of the deep-underground repository in 2015, to safely dispose of spent nuclear fuel. Up to 6,500 tonnes of radioactive uranium will be packed in copper canisters, which will be covered with clay and buried inside tunnels of granite bedrock 400 metres underground. The nuclear material will remain sealed off there for several hundred thousand years — by which time the radiation levels will be harmless.

Nature **613**, 11-12 (2023)

doi: https://doi.org/10.1038/d41586-022-04444-3

References

1. Van Dyck, C. H. *et al. N. Engl. J. Med.* https://doi.org/10.1056/NEJMoa2212948 (2022).

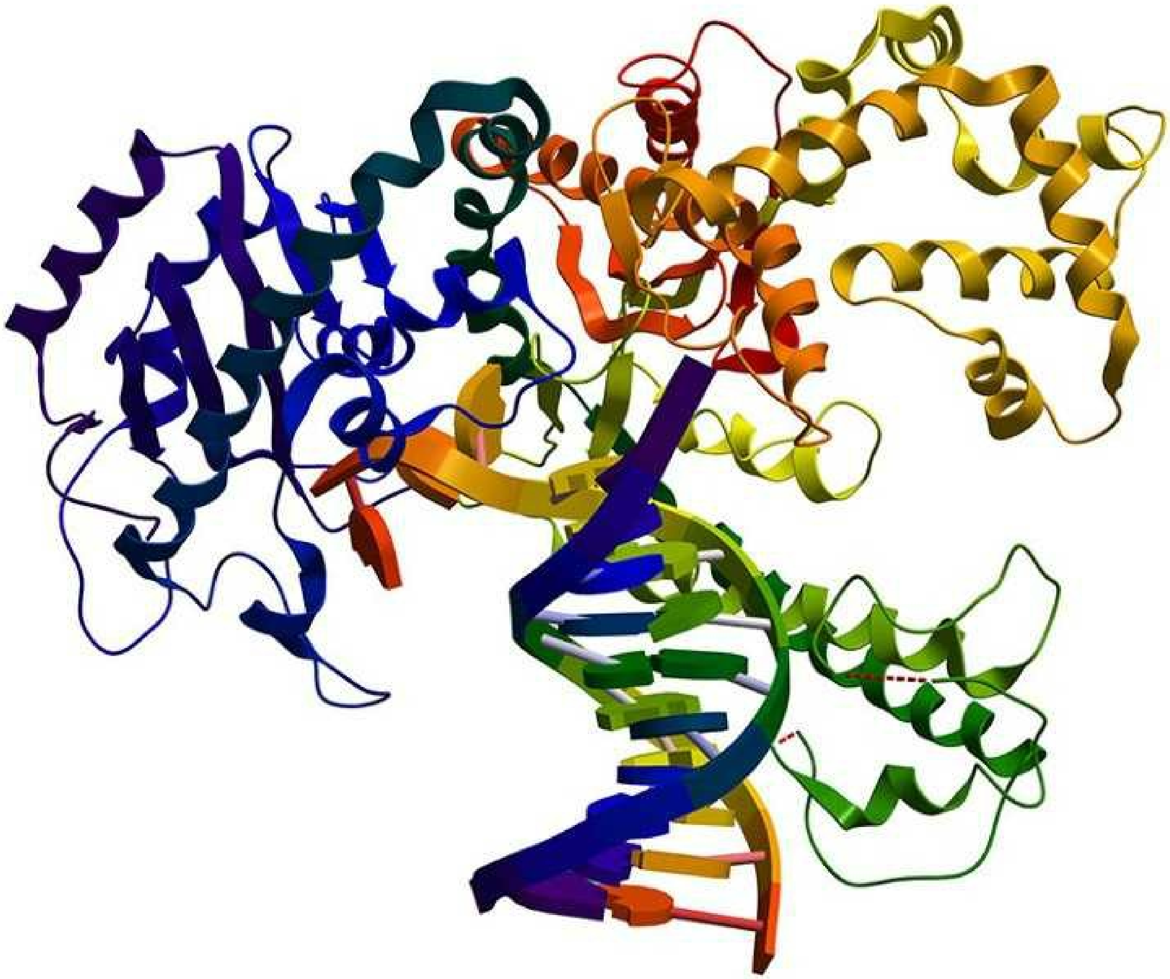
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-022-04444-3

- NEWS
- 13 December 2022

After AlphaFold: protein-folding contest seeks next big breakthrough

Two years after DeepMind's revolutionary AI swept a competition for predicting protein structures, researchers are building on AlphaFold's success.

- Ewen Callaway



A protein’s function is determined by its 3D shape.Credit: Leonid Andronov/Alamy

“In some sense the problem is solved,” computational biologist John Moult declared in late 2020. The London-based company DeepMind had just swept a biennial contest co-founded by Moult that tests teams’ abilities to predict protein structures — one of biology’s grandest challenges — with its revolutionary artificial-intelligence (AI) tool AlphaFold.

Two years later, Moult’s competition, the Critical Assessment of Structure Prediction (CASP), is still walking in AlphaFold’s long shadow. Results from this year’s edition (CASP15) — which were unveiled over the weekend at a conference in Antalya, Turkey — show that the most successful approaches

to predicting protein structures from their amino-acid sequences incorporated AlphaFold, which relies on an AI approach called deep learning. “Everyone is using AlphaFold,” says Yang Zhang, a computational biologist at the University of Michigan in Ann Arbor.

Yet AlphaFold’s progress has opened the floodgates for new challenges in protein-structure prediction — some included in this year’s CASP — that might require new approaches and more time to fully tackle. “The low-hanging fruit has been picked,” says Mohammed AlQuraishi, a computational biologist at Columbia University in New York City. “Some of the next problems are going to be harder.”

Matchmaking

CASP started in 1994, aiming to bring rigour to the field of protein-structure prediction — progress on which would accelerate efforts to understand the building blocks of cells and advance drug discovery. During the year of a contest, teams are tasked with using computational tools to predict the structures of proteins that have been determined using experimental methods such as X-ray crystallography and cryo-electron microscopy, but not yet released.

Entries are assessed according to how well predictions for entire proteins, or independently folding subunits called domains, match the experimental structures. Some of AlphaFold’s predictions at CASP14 were more or less indistinguishable from the experimental models — the first time such accuracy had been achieved.

Since its unveiling at CASP14, AlphaFold has become omnipresent in life-sciences research. DeepMind released the software’s underlying code in 2021 so that anyone could run the program, and an AlphaFold database updated this year holds predicted structures — of varying quality — for almost every protein from all organisms represented in genome databases, a total of more than 200 million proteins.

AlphaFold’s success and newfound ubiquity presented a challenge to Moult, who is at the University of Maryland, Rockville, and his colleagues as they

planned this year’s CASP. “People say, ‘Oh, we don’t need CASP anymore, the problem was solved.’ And I think that’s exactly the wrong way round.”

At CASP15, the most successful teams were those that had adapted and built on AlphaFold in various ways, leading to modest gains in predicting the shape of individual proteins and domains. “The accuracy is already so high that it’s hard to get much better,” says Moult.

Protein complexes

To make the competition more relevant in a post-AlphaFold world, Moult and his team added new challenges and tweaked some existing ones. New tests include determining how proteins interact with other molecules such as drugs and predicting the multiple shapes that some proteins can assume. For the past decade, CASP has included ‘complexes’ of multiple interacting proteins, says Moult, but accurately predicting the structure of such molecules has taken on added emphasis this year.

“That is the right thing to do,” says Zhang, because predicting the structures of single proteins or domains — the bread and butter of past CASPs — has largely been solved by AlphaFold. Determining the shape of protein complexes, in particular, represents an important new challenge for the field, because there is a lot of room for improvement, says Arne Elofsson, a protein bioinformatician at Stockholm University.

AlphaFold was initially designed to predict the shape of individual proteins. But, within days of its public release, other scientists showed that the software could be ‘hacked’ to model how multiple proteins interact. In the months since then, researchers have come up with myriad approaches to improve AlphaFold’s ability to tackle complexes. DeepMind even released an update called AlphaFold-Multimer with that goal in mind.

Such efforts seem to have paid off, because CASP15 saw a marked increase in the number of accurate complexes, compared with previous contests, mainly due to methods that adapted AlphaFold. “It’s a new game for us to be close to experimental accuracy with complexes,” says Moult. “We’ve got some failures too.”

For instance, teams made stunningly accurate predictions of a viral molecule of unknown function made up of two identical intertwined proteins. This kind of shape bamboozled pre-AlphaFold tools, says Ezgi Karaca, a computational structural biologist at the Izmir Biomedicine and Genome Center in Turkey, who assessed the complex predictions. The standard version of AlphaFold failed to accurately model the shape of a giant, 20-chain bacterial enzyme, but some teams predicted the protein's structure by applying extra hacks to the network, Karaca adds.

Meanwhile, teams struggled to predict complexes involving immune molecules called antibodies — including several attached to a SARS-CoV-2 protein — and related molecules called nanobodies. But there were glimpses of success in some teams' predictions, says Karaca, suggesting that hacks to AlphaFold will be useful for predicting the shape of these medically important molecules.

Time out

This year's CASP was also notable for the absence of DeepMind. The company did not state its reason for not participating, but released a short statement during CASP15 congratulating the teams that did. (At the same time, it rolled out an update to AlphaFold to help researchers benchmark their progress against the network.)

Other researchers say the competition is a considerable time commitment, which the company might have felt was better spent on other challenges. "It would have been nice for us if they had participated," Moult says. But he adds that "because the methods are so good, they couldn't do another big leap".

Making big improvements to AlphaFold will take time, say researchers, and will probably require new innovations in machine learning and protein-structure prediction. One area under development is the application of 'language models', such as those used in predictive-text tools, to the prediction of protein structures. But these methods — including one developed by the social-networking giant Meta — did not perform nearly as well at CASP15 as did tools based on AlphaFold.

Such tools might, however, be useful for predicting how mutations alter a protein's structure — one of several key challenges in protein-structure prediction that has emerged as a result of AlphaFold's success. Thanks to this, the field is no longer focused on a single goal, AlQuraishi says. "There's a whole slew of these problems."

Nature **613**, 13-14 (2023)

doi: https://doi.org/10.1038/d41586-022-04438-1

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-022-04438-1

- NEWS EXPLAINER
- 13 December 2022

Are we in the Anthropocene? Geologists could define new epoch for Earth

Researchers have zeroed in on nine sites that could describe a new geological time, marked by pollution and other signs of human activity.

- McKenzie Prillaman



According to some geologists, the Anthropocene epoch is defined by markers

of human activity — including fossil-fuel emissions — that have altered Earth. Credit: Jochen Tack/Alamy

Geologists could soon decide which spot on Earth marks the first clear evidence of the Anthropocene — which many of them think is a new geological epoch that began when humans started altering the planet with various forms of industrial and radioactive materials in the 1950s. They have so far whittled their choices down to nine candidate sites worldwide (see 'Defining the Anthropocene'), each being considered for how reliably its layers of mud, ice or other matter tell the story of people's influence on a timeline that extends billions of years into the past.



Humans versus Earth: the quest to define the Anthropocene

If the nearly two dozen voting members of the Anthropocene Working Group (AWG), a committee of scientists formed by the International Commission on Stratigraphy (ICS), agree on a site, the decision could usher in the end of the roughly 12,000-year-old Holocene epoch. And it would officially acknowledge that humans have had a profound influence on Earth.

"We're pointing to something in the rock record that shows we've changed the planet," says Kristine DeLong, a palaeoclimatologist at Louisiana State University in Baton Rouge who studies the West Flower Garden Bank, a candidate site in the Gulf of Mexico.

The Anthropocene site will join 79 others that physically define stages of Earth’s geological timescale — that is, if it’s approved. Even if the AWG agrees on a final candidate, several other committees of geologists must vote on the selection before it is made official. And not all scientists agree that it should be.

Here, *Nature* examines what it will take to formally define the Anthropocene epoch.

Why do some geologists want an Anthropocene marker?

Scientists coined the term Anthropocene in 2000, and researchers from several fields now use it informally to refer to the current geological time interval, in which human activity is driving Earth’s conditions and processes. Formalizing the Anthropocene would unite efforts to study people’s influence on Earth’s systems, in fields including climatology and geology, researchers say. Transitioning to a new epoch might also coax policymakers to take into account the impact of humans on the environment during decision-making.



Coral grows on an oil rig in Flower Garden Banks National Marine Sanctuary, in the Gulf of Mexico.Credit: Flip Nicklin/Minden Pictures/Alamy

“It’s a label,” says Colin Waters, who chairs the AWG and is a geologist at the University of Leicester, UK. “It’s a great way of summarizing a lot of concepts into one word.”

Mentioning the Jurassic period, for instance, helps scientists to picture plants and animals that were alive during that time, he says. “The Anthropocene represents an umbrella for all of these different changes that humans have made to the planet,” he adds.

How do scientists usually choose sites that define the geological timeline?

Typically, researchers will agree that a specific change in Earth’s geology

must be captured in the official timeline. The ICS will then determine which set of rock layers, called strata, best illustrates that change, and it will choose which layer marks its lower boundary. This is called the Global Stratotype Section and Point (GSSP), and it is defined by a signal, such as the first appearance of a fossil species, trapped in the rock, mud or other material. One location is chosen to represent the boundary, and researchers mark this site physically with a golden spike, to commemorate it.

Defining the Anthropocene: nine sites are in the running to be given the ‘golden spike’ designation

Site	Location	Material type	How it captures signs of human activity
Beppu Bay	Kyushu Island, Japan	Marine sediment	Sediment fell to the sea floor, where the oxygen-deprived water limits disturbance by animals.
Crawford Lake	Ontario, Canada	Lake mud	Particles accumulated at the steep-sloped lake bottom, where water layers don’t mix.
Flinders Reef	Coral Sea, Australia	Coral	The corals’ growing exoskeletons have trapped chemicals and particles.
Gotland Basin	Baltic Sea	Marine sediment	Sediment fell to the sea floor, where the oxygen-deprived water limits disturbance by animals.

Palmer ice core	Antarctic Peninsula	Ice	Annual layers of snowfall captured particles and chemicals from the air.
Searsville Lake	California	Lake mud	Silt layers accumulated at the bottom after storm events.
Sihailongwan Lake	Jilin province, China	Lake mud	Particles fell to the bottom of this lake, which has an oxygen-deprived lower layer that limits disturbance.
Śnieżka peat bog	Sudetes Mountains, Poland	Peat layer	Peat in this high-altitude bog has captured chemicals and particles from the air.
West Flower Garden Bank	Gulf of Mexico	Coral	The corals' growing exoskeletons have trapped chemicals and particles.

Source: Anthropocene Working Group

But the Anthropocene has posed problems. Geologists want to capture it in the timeline, but its beginning isn't obvious in Earth's strata, and signs of human activity have never before been part of the defining process. The AWG was established in 2009 to explore whether the Anthropocene should enter the geological timescale and, if so, how to define its start.

"We were starting from scratch," says Jan Zalasiewicz, a geologist at the University of Leicester who formerly chaired the AWG and remains a voting member. "We had a vague idea about what it might be, [but] we didn't know what kind of hard evidence would go into it."

Years of debate among the group's multidisciplinary members led them to identify a host of signals — radioactive isotopes from nuclear-bomb tests, ash from fossil-fuel combustion, microplastics, pesticides — that would be

trapped in the strata of an Anthropocene-defining site. These began to appear in the early 1950s, when a booming human population started consuming materials and creating new ones faster than ever.



This golden spike in the Flinders Ranges of South Australia was approved by geologists in 2004, to mark strata exemplifying the Ediacaran period.Credit: James St. John (CC BY 2.0)

During a review that took place a few months ago, the AWG narrowed its list from 12 to 9 candidate sites, tossing out certain locations because their layers weren't ideal. Among the sites remaining is Crawford Lake in Ontario, Canada, which is described as a sinkhole by Francine McCarthy, a geologist at Brock University in St Catharines, Canada, who studies the location. "The lake itself isn't very big in area, but it's very, very deep," she says. Particles that fall into the lake settle at the bottom and accumulate into undisturbed layers.

Another site on the shortlist is West Flower Garden Bank. Corals here could become a living golden spike because they constantly build new exoskeletons

that capture chemicals and particles from the water, DeLong says. “The skeleton has layers in it, kind of like tree rings,” she adds.

Why do some geologists oppose the Anthropocene as a new epoch?

“It misrepresents what we do” in the ICS, says Stanley Finney, a stratigrapher at California State University, Long Beach, and secretary-general for the International Union of Geological Sciences (IUGS). The AWG is working backwards, Finney says: normally, geologists identify strata that should enter the geological timescale before considering a golden spike; in this case, they’re seeking out the lower boundary of an undefined set of geological layers.



Involve social scientists in defining the Anthropocene

Lucy Edwards, a palaeontologist who retired in 2008 from the Florence Bascom Geoscience Center in Reston, Virginia, agrees. For her, the strata that might define the Anthropocene do not yet exist because the proposed epoch is so young. “There is no geologic record of tomorrow,” she says.

Edwards, Finney and other researchers have instead proposed calling the Anthropocene a geological ‘event’, a flexible term that can stretch in time,

depending on human impact. “It’s all-encompassing,” Edwards says.

Zalasiewicz disagrees. “The word ‘event’ has been used and stretched to mean all kinds of things,” he says. “So simply calling something an event doesn’t give it any wider meaning.”

What happens next?

In a recent Perspective article in *Science*, Waters and AWG secretary Simon Turner at University College London wrote that the committee would vote to choose a single site by the end of this year[1]. But 60% of the group’s voting members must agree on a final candidate — and, with several sites under consideration, Waters isn’t sure that a consensus can be reached anytime soon. If no clear winner emerges this month, more voting will be needed to narrow the candidate list, delaying a decision possibly until May 2023.



Anthropocene now: influential panel votes to recognize Earth’s new epoch

And that’s not the end of the process. After selecting a finalist, the AWG will present its findings to the ICS’s Subcommission on Quaternary Stratigraphy. Favourable votes from this group would move the proposal to another ICS committee, and subsequent approval would push it to the final stage: ratification by the IUGS.

But the motion could fail at any of those points. And if it does, the AWG will have to revamp its proposal before it can try again — and possibly nominate a new golden-spike site.

Regardless of the outcome, Zalasiewicz thinks that the AWG's work to define the Anthropocene has been useful. What everybody wants to know is how humans are changing the planet's geology, he says. "That is the underlying reality that we're trying to describe."

Nature **613**, 14-15 (2023)

doi: https://doi.org/10.1038/d41586-022-04428-3

References

1. Waters, C. N. & Turner, S. D. *Science* **378**, 706–708 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-022-04428-3

- NEWS FEATURE
- 04 January 2023

The ‘breakthrough’ obesity drugs that have stunned researchers

A class of drugs that quash hunger have shown striking results in trials and in practice. But can they help all people with obesity — and conquer weight stigma?

- McKenzie Prillaman



Illustration by Karol Banach

The hotel ballroom was packed to near capacity with scientists when Susan

Yanovski arrived. Despite being 10 minutes early, she had to manoeuvre her way to one of the few empty seats near the back. The audience at the ObesityWeek conference in San Diego, California, in November 2022, was waiting to hear the results of a hotly anticipated drug trial.

The presenters — researchers affiliated with pharmaceutical company Novo Nordisk, based in Bagsværd, Denmark — did not disappoint. They described the details of an investigation of a promising anti-obesity medication in teenagers, a group that is notoriously resistant to such treatment. The results astonished researchers: a weekly injection for almost 16 months, along with some lifestyle changes, reduced body weight by at least 20% in more than one-third of the participants[1]. Previous studies[2,3] had shown that the drug, semaglutide, was just as impressive in adults.

The presentation concluded like no other at the conference, says Yanovski, co-director of the Office of Obesity Research at the US National Institute of Diabetes and Digestive and Kidney Diseases in Bethesda, Maryland. Sustained applause echoed through the room “like you were at a Broadway show”, she says.

This energy has pervaded the field of obesity medicine for the past few years. After decades of work, researchers are finally seeing signs of success: a new generation of anti-obesity medications that drastically diminish weight without the serious side effects that have plagued previous efforts.

These drugs are arriving in an era in which obesity is growing exponentially. Worldwide obesity has tripled since 1975; in 2016, about 40% of adults were considered overweight and 13% had obesity, according to the World Health Organization (WHO). With extra weight often comes heightened risk of health conditions such as type 2 diabetes, heart disease and certain cancers. The WHO recommends healthier diets and physical activity to reduce obesity, but medication might help when lifestyle changes aren’t enough. The new drugs mimic hormones known as incretins, which lower blood sugar and curb appetite. Some have already been approved for treating type 2 diabetes, and they are starting to win approval for inducing weight loss.



[Why are obese people more vulnerable to COVID?]

The ability to melt weight away by tweaking biology gives credence to the idea that obesity is a disease. In the past, scientists and the public often thought that those with obesity simply lacked the willpower to lose weight. But evidence is growing that most people's bodies have a natural size that can be hard to change. "The body will defend its weight," says Richard DiMarchi, a chemist at Indiana University Bloomington.

However, some researchers worry that these drugs play into some societies' obsession with being thin. Body size isn't always a good predictor of health. "I'm really hesitant to be excited about something that I think is potentially harmful from a weight stigma perspective," says Sarah Nutter, a psychologist at the University of Victoria in Canada, who specializes in weight stigma and body image.

Research questions abound, including who will respond to treatment and whether people will have to take these drugs for life — a huge barrier to access, given that they also carry a hefty price tag: the injections often cost upwards of US$1,000 each month.

Still, obesity researchers are celebrating these developments. For the first time, scientists can pharmacologically alter weight safely, says physician-scientist Matthias Tschöp, chief executive of Helmholtz Munich in Germany.

"It indeed is 'the' transformative breakthrough."

Hormone hunt

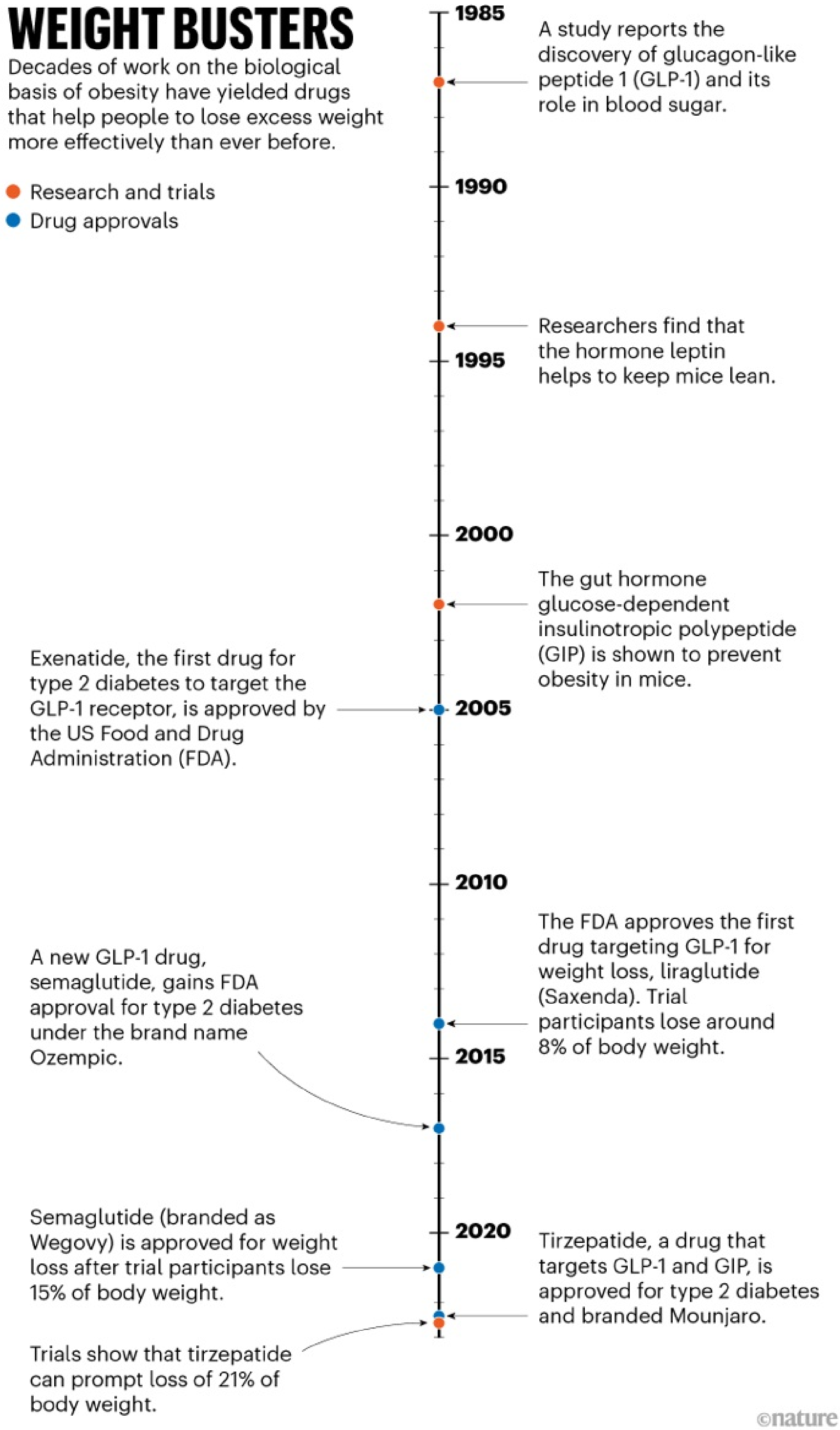
The seeds of today's success were sown decades ago, when Jeffrey Friedman was racing to figure out which gene mutation was making the mice in his laboratory eat until they became obese. In 1994, Friedman, a molecular geneticist at The Rockefeller University in New York City, discovered that the faulty gene encoded leptin, a hormone that is produced by fat tissue and induces a feeling of fullness[4]. Giving leptin supplements to mice that lacked it reduced their hunger and body weight.

"That really revolutionized our thinking about the biological basis of obesity and appetite regulation," Yanovski says.

An explosion of research into obesity's underpinnings followed, alongside research into pharmacological treatments. But these early drugs led to only modest weight loss and serious side effects, especially on the heart.

Even before leptin's discovery, researchers had been looking for hormones that regulate blood glucose levels, and had found one called GLP-1 (glucagon-like peptide 1). It seemed to have the opposite effect of type 2 diabetes — GLP-1 enhanced insulin production and reduced blood sugar[5] — making it an appealing approach to treating obesity, says Jens Juul Holst, a medical physiologist at the University of Copenhagen, who discovered and characterized GLP-1.

In the 2000s, the US Food and Drug Administration (FDA) began approving drugs that mimicked GLP-1 as type 2 diabetes treatments (see 'Weight busters'). But scientists noticed that participants in clinical trials also lost weight, owing to GLP-1's effect on receptors in the brain that govern appetite and those in the gut that slow digestion. Over time, companies began to trial these diabetes medications for weight loss. By the mid-2010s, one such drug, liraglutide, was capable of eliciting a loss in body weight of about 8% on average, 5 percentage points more than for people taking a placebo[6] — clinically relevant, but not astonishing.


WEIGHT BUSTERS
Decades of work on the biological basis of obesity have yielded drugs that help people to lose excess weight more effectively than ever before.
Research and trials
Drug approvals
1985
1990
1995
2000
2005
2010
2015
2020
A study reports the discovery of glucagon-like peptide 1 (GLP-1) and its role in blood sugar.
Researchers find that the hormone leptin helps to keep mice lean.
The gut hormone glucose-dependent insulinotropic polypeptide (GIP) is shown to prevent obesity in mice.
Exenatide, the first drug for type 2 diabetes to target the GLP-1 receptor, is approved by the US Food and Drug Administration (FDA).
The FDA approves the first drug targeting GLP-1 for weight loss, liraglutide (Saxenda). Trial participants lose around 8% of body weight.
A new GLP-1 drug, semaglutide, gains FDA approval for type 2 diabetes under the brand name Ozempic.
Semaglutide (branded as Wegovy) is approved for weight loss after trial participants lose 15% of body weight.
Tirzepatide, a drug that targets GLP-1 and GIP, is approved for type 2 diabetes and branded Mounjaro.
Trials show that tirzepatide can prompt loss of 21% of body weight.
©nature

But in early 2021, scientists were wowed by a phase III clinical trial[2] investigating a new drug of the same type: semaglutide. The molecule, a modified version of liraglutide, acts on the same pathways but remains intact and active in the body for longer, says DiMarchi. It might also have better access to brain regions that regulate appetite, he adds.

Those receiving weekly injections of semaglutide lost, on average, 14.9% of their body weight after 16 months of treatment; those who received a placebo lost 2.4% on average. In 2021, four years after approving it for diabetes, the FDA approved semaglutide for weight loss for adults with obesity.

Historically, it hasn't been possible to safely decrease body weight by more than 10% through pharmacological methods, says Timo Müller, a biologist and director of the Helmholtz Munich Institute for Diabetes and Obesity. But these newer treatments also improve cardiovascular health, he adds — the opposite of past iterations.

There could now be an even more effective drug in town: tirzepatide. Tirzepatide doesn't just target the GLP-1 receptor; it also mimics another hormone involved in insulin secretion, known as glucose-dependent insulinotropic polypeptide (GIP). Approved in 2022 for type 2 diabetes, this treatment — developed by Eli Lilly, based in Indianapolis, Indiana — led to a 21% drop in body weight, on average, at the highest dose, compared with 3% for placebo[7].



[How obesity could create problems for a COVID vaccine]

It's unclear why mimicking both hormones works better than imitating just one. Müller says that tirzepatide might be a more potent activator of the GLP-1 receptor, and that GIP might help to make GLP-1's side effects more tolerable, allowing for higher doses. It's also possible that GIP might drive some weight loss on its own.

Despite the uncertainties, the levels of weight loss following tirzepatide treatment approach those typically achievable only through bariatric surgery. This procedure reduces body weight by 30% or more after six months, and the weight loss continues for the next year or two.

"Ten years ago, if you had told me we have something that gets us pretty close [to bariatric surgery], I would have said that's not possible," says Ruth Gimeno, group vice-president of diabetes, obesity and cardiometabolic research and early clinical development at Eli Lilly. The company plans to apply for the drug to be approved, pending results from a second phase III trial wrapping up in April 2023.

Mechanism mystery

Despite tirzepatide's promising results, it has researchers puzzled. It's clear how GLP-1 helps to spur weight loss, but GIP's role is a surprise. In fact, scientists have long thought that GIP actually encourages obesity: mice with dysfunctional GIP receptors are resistant to obesity[8]. Therefore, to induce weight loss, researchers thought the receptor should be switched off. But tirzepatide does the opposite.

"We were the first who came up with this crazy idea," says Müller, who collaborates with Novo Nordisk. "And we were quite heavily criticized in the field."

Müller and his colleagues — including DiMarchi and Tschöp — knew that GIP stimulates insulin secretion depending on blood glucose levels, just like GLP-1, says Müller. So they developed molecules that mimicked both hormones. After initial studies demonstrated that activating both the GIP and

GLP-1 receptors caused weight loss, pharmaceutical companies created their own molecules achieving the same results, thus confirming that the method worked.



A doctor weighs a woman in Spain as part of a weight loss challenge. Diet and lifestyle changes are recommended approaches to weight loss, but a new class of drugs could help. Credit: Miguel Riopa/AFP/Getty

However, not everyone has changed their views on GIP. Holst feels that tirzepatide is simply a super powerful GLP-1 imitator.

It can also mimic GIP, “but it doesn’t really matter in patients with diabetes and obesity, because the GIP part doesn’t really do anything,” says Holst. Eli Lilly is conducting early-stage clinical trials with drugs that target GIP alone, which Holst says will resolve the ongoing debate.

And biopharmaceutical company Amgen, based in Thousand Oaks, California, is pursuing a drug that activates the GLP-1 receptor while

thwarting the GIP receptor. Early clinical-trial data show that this treatment reduced body weight by up to around 15% after 12 weeks.

Other approaches include ‘triple agonists’ that mimic the actions of GLP-1, GIP and a third hormone, glucagon, which also stimulates insulin secretion[9]. Still other gut hormones involved in appetite, such as peptide YY, are being explored, too. And some researchers are investigating the monoclonal antibody bimagrumab, which increases muscle mass while decreasing fat.

Open questions

One big question facing researchers now is whether people will need to take these medications for life to maintain their weight. A subset of clinical-trial participants who ceased taking semaglutide and stopped the study’s lifestyle interventions regained about two-thirds of their lost weight after one year[10].

Another unknown is who will respond to these drugs — and who won’t. It’s too early to tell now, but the drugs seem to be less effective for weight loss in people with type 2 diabetes than in those without. Conditions such as fatty liver disease and having fat around the organs, known as visceral body fat, might also affect how people respond to different drugs, Tschöp says.

Some researchers also worry that by offering a weight solution in societies that prize thinness, these drugs could also inadvertently reinforce the disputed link between excess weight and health. One study found that nearly 30% of people who are considered obese are metabolically healthy[11]. Another showed that other health problems tend to be a better predictor of someone’s risk of death than is weight[12], demonstrating the need to consider factors other than weight when judging health, says Nutter.

“To pathologize a person’s health simply based on their body weight is potentially really, really harmful,” she adds.

Nutter is concerned that people might start these treatments — whose side effects, such as nausea and vomiting, can be severe — to escape weight stigma, rather than to serve a true health need.

Others worry about the idea that these drugs offer a quick fix. This is a common misconception about bariatric surgery, says Leslie Heinberg, a clinical psychologist at the Cleveland Clinic in Ohio who specializes in bariatric behavioural health and body image. “Some people who still hold on to those mistaken beliefs will say, ‘Oh, now people can just take this pill and that’s the easy way out of obesity,’” she says.

Still, there is plenty of demand. And although these drugs are entering the market, not everyone who needs them will have access.

For a start, they are pricey — semaglutide for weight loss, branded as Wegovy, costs about $1,300 a month — and many insurance companies in the United States refuse to cover the expense, primarily owing to a misunderstanding of what causes obesity and viewing the treatments as ‘vanity drugs’.

“People talk about some of these drugs as being game-changers,” says Patty Nece, chair of the board of directors of the Obesity Action Coalition (OAC), an advocacy group based in Tampa, Florida. But, she adds, “for an individual patient, it’s never going to be a game-changer if they can’t afford it or don’t get access to it”.

Organizations such as the OAC are pushing pharmaceutical companies to offer affordability programmes. Eli Lilly, for example, has a ‘bridging programme’ for Mounjaro — tirzepatide for type 2 diabetes — under which the medication can cost as little as $25 for the first three months. Novo Nordisk has a similar programme for Wegovy.

Whatever the upfront costs, some scientists stress that addressing obesity could allow health-care systems to save enormous amounts of money by reducing a slew of conditions that are linked to the disease.

Although researchers are still chipping away at obesity’s complex combination of causes — including genetics, environment and behaviour — many support the idea that biology plays a significant part. Eating healthily and exercising will always be part of treatment, but many think that these drugs are a promising add-on. And some researchers think that because these drugs act through biological mechanisms, they will help people to understand

that a person’s body weight is often beyond their control through lifestyle changes alone. “Tirzepatide very clearly shows that it’s not about willpower,” Gimeno says.

Nature **613**, 16-18 (2023)

doi: https://doi.org/10.1038/d41586-022-04505-7

References

1. Weghuber, D. *et al. N. Engl. J. Med.* **387**, 2245–2257 (2022).
2. Wilding, J. P. H. *et al. N. Engl. J. Med.* **384**, 989–1002 (2021).
3. Davies, M. *et al. Lancet* **397**, 971–984 (2021).
4. Zhang, Y. *et al. Nature* **372**, 425–432 (1994).
5. Holst, J. J., Ørskov, C., Vagn Nielsen, O. & Schwartz, T. W. *FEBS Lett.* **211**, 169–174 (1987).
6. Pi-Sunyer, X. *et al. N. Engl. J. Med.* **373**, 11–22 (2015).
7. Jastreboff, A. M. *et al. N. Engl. J. Med.* **387**, 205–216 (2022).
8. Miyawaki, K. *et al. Nature Med.* **8**, 738–742 (2002).
9. Urva, S. *et al. Lancet* **400**, 1869–1881 (2022).
10. Wilding, J. P. H. *et al. Diabetes Obes. Metab.* **24**, 1553–1564 (2022).
11. Tomiyama, A., Hunger, J., Nguyen-Cuu, J. & Wells, C. *Int. J. Obes.* **40**, 883–886 (2016).
12. Li, J. *et al. PLoS ONE* **16**, e0253696 (2021).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-022-04505-7

| Section menu | Main menu |

- NEWS FEATURE
- 20 December 2022

Twitter changed science — what happens now it's in turmoil?

The microblogging platform has transformed research communication, but its future is in doubt.

- Chris Stokel-Walker



Credit: NurPhoto/Getty

In November, Vince Knight decided he'd had enough of Twitter. After more

than a decade on the social-media platform, Knight — a mathematician at Cardiff University, UK — was concerned about the site's direction under its new owner, entrepreneur Elon Musk, who began laying off vast numbers of staff shortly after he acquired it. "Twitter is getting uncomfortable," wrote Knight on the platform; he then jumped ship to Mastodon, a competing service. He says he simply didn't want to support Musk's Twitter any more.

The past few weeks have been tumultuous for Twitter. After Musk laid off staff, the site has repeatedly malfunctioned as the remaining engineers have struggled to keep on top of issues. Musk has also said he wants to take the platform in a new direction, encouraging accounts that were previously banned to return. Some reports, including one from researchers at Tufts University in Medford, Massachusetts, say abuse is rising on the platform (see go.nature.com/3vcgpfw).

On 11 December, Musk tweeted that his "pronouns are Prosecute/Fauci" in an apparent attempt both to mock the transgender and gender-nonconforming rights movements and to malign the departing director of the US National Institute of Allergy and Infectious Diseases, Anthony Fauci, who has faced abuse and death threats for his role in advising the US government response to COVID-19.

Musk's erratic and confrontational management of Twitter has worried many users, including researchers such as Knight. For hundreds of thousands of scientists, Twitter is a sounding board, megaphone and common room: a place to broadcast research findings, debate issues in academia and interact with people who they wouldn't normally meet up with.



Should I join Mastodon? A scientists' guide to Twitter's rival

"I would never be able to know so many scientists without it," says Oded Rechavi, who works on transgenerational inheritance at Tel Aviv University in Israel. "It increases democracy in science and gives you more opportunities, no matter where you are."

Since the site's founding in 2006, Twitter executives have often asserted that it aims to be nothing less than a 'public town square' of communication; it now claims almost 250 million daily users. At that scale, abuse, misinformation and bots have been ever-present, but for many researchers, the advantages of rapid, widespread communication to each other and an engaged public outweighed these problems.

The threat of Twitter changing radically under its new management, or perhaps disappearing altogether, has raised concerns and questions for researchers. How well has this vast social-media platform benefited science, and to what extent has it harmed it? If it disappears, would researchers want to recreate it elsewhere?

Twitter's influence on science

No one knows how many researchers have joined Twitter, but this August,

Rodrigo Costas Comesana, an information scientist at Leiden University in the Netherlands, and his colleagues published a data set of half a million Twitter users[1] who are probably researchers. (The team used software to try to match details of Twitter profiles to those of authors on scientific papers.) In a similar, smaller 2020 study, Costas and others estimated that at least 1% of paper authors in the Web of Science had profiles on Twitter, with the proportion varying by country[2]. A 2014 *Nature* survey found that 13% of researchers used Twitter regularly, although respondents were mostly English-speaking and there would have been self-selection bias (see *Nature* **512**, 126–129; 2014).



How Musk's takeover might change Twitter: what researchers think

Even though many researchers aren't on Twitter, the platform has a major role in science communication, according to several studies. "Typically, about one-third of all the scientific literature gets tweeted," says Costas, pointing to a 2020 study[3] that analysed 12 million papers from 2012–18; by 2018, the proportion tweeted had nearly doubled from 2012 levels, to almost 40%. And during the SARS-CoV-2 pandemic, more than half of all journal articles on COVID-19 published up to April 2021 were mentioned at least once on Twitter[4].

All of this tweeting hasn't necessarily led to engagement, however: a pre-pandemic study by Costas and his team analysed 1.1 million links to scholarly articles posted on Twitter up to September 2019. They found that

half of those posts drew no clicks to the underlying research, whereas 22% received only one or two clicks[5].

But for many scientists, Twitter has become an essential tool for collaboration and discovery — a source of real-time conversations around research papers, conference talks and wider topics in academia. Papers now zip around scientific communities faster thanks to Twitter, says Johann Unger, a linguist at Lancaster University, UK, who notes that extra information is also shared in direct private messages through the site. And its limit on tweet length — currently 280 characters — has pushed academics into keeping their commentary pithy, he adds.

The social platform has flattened hierarchies, throwing people into conversations regardless of geography, seniority or specialism. “Academia is characterized by a lot of gatekeeping,” says Daniel Quintana, a psychologist at the University of Oslo, who has written an e-book on how scientists can use Twitter (https://t4scientists.com). “Twitter provides a fantastic way to actually get your work out there.”

It has also given an influential voice to people who might otherwise be excluded, and has helped to broker support networks for those who don’t see people like them in their own departments, says Sigourney Bonner, co-founder of the #BlackinCancer community and a PhD student at Cancer Research UK’s Cambridge Institute. “I didn’t meet a Black woman with a PhD until I started my own,” she says. Movements united by hashtags — from #IAmAScientistBecause to #BlackInTheIvory — have often seen Twitter acting as a rallying point for discussing key problems in academia, such as racism, sexism, harassment and bullying.

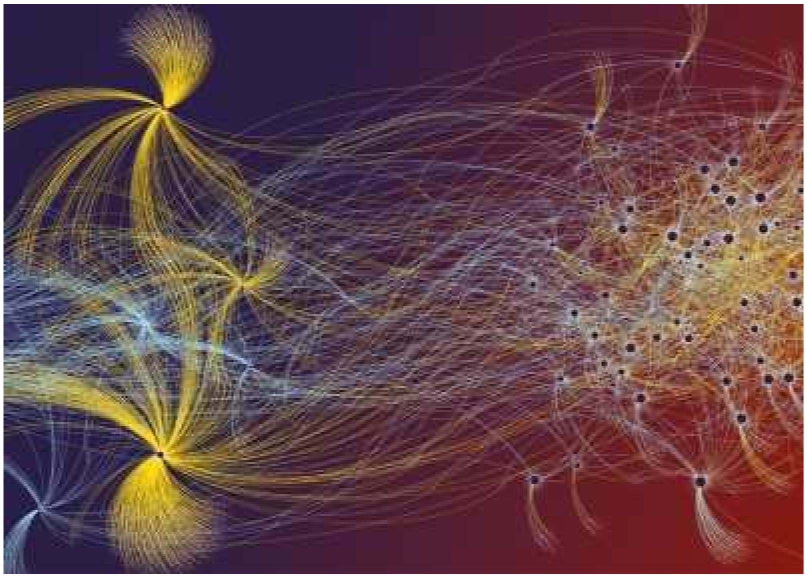


Sigourney Bonner, who co-founded the #BlackinCancer community.Credit: Cancer Research UK

Because of its status as a pre-eminent public discussion network and its relatively open data, Twitter has become a hotbed for researchers studying social reactions to world events — in particular, how information spreads on the network. A *Nature* analysis of the Scopus database of scientific literature, for this article, found more than 41,000 articles and conference papers that mention Twitter in the title, abstract or keywords. That number has increased from just one in 2006 to more than 4,800 in 2022.

In a widely shared study from 2018, researchers at the Massachusetts Institute of Technology in Cambridge studied Twitter and found that false news stories on the site spread much faster than do true news stories — possibly because, they reported, the false news items had more 'novelty' than the true news[6]. The false news also tended to arouse emotions such as fear, disgust and surprise.

And in a 2018 study of hate speech on Twitter, Manoel Horta Ribeiro, now a PhD student at the Swiss Federal Institute of Technology (EPFL) in Lausanne, and his colleagues found that users whose tweets contained hate speech tweeted more often than those who didn't use such language, and were retweeted more frequently than their less-incendiary counterparts[7].



[How Facebook, Twitter and other data troves are revolutionizing social science](#)

These studies and more point to the conundrums that Twitter poses for scientists and other users. Like other platforms that are financed mostly by advertisements, Twitter aims largely to retain people's engagement and attention. Accordingly, Twitter's discovery algorithm (which surfaces heavily discussed or shared messages on people's timelines) "prioritizes a very particular type of content", says Renée DiResta, who studies social networks and misinformation at the Stanford Internet Observatory in California. "People who maybe don't necessarily have an institutional credential, but are adept at commenting on a particular topic, can capture public attention," she says.

The idea of Twitter as a great democratizer also doesn't always match reality, DiResta adds. Accounts with a large, established following have much greater reach than "your average science experts on the platform", she says.

And although Twitter's algorithms elevate humour, delight and entertainment, they can also encourage performative tweets, dismissive arguments and snide comments that veer into abuse. Real-time criticism can swiftly turn ugly, and users can easily butt in on others' conversations, with hordes of people sometimes exhorted to insult and mock a specific target.

Twitter has always struggled to cope with how to moderate such rapid shifts in online conversation. It's a problem that seems likely to worsen now that Musk has made cuts to the company's staff and its safety systems.

Pandemic Twitter

This double-edged nature of Twitter has never been clearer than during the COVID-19 pandemic. Many academics built up large public followings through their expert analyses about SARS-CoV-2, and made fruitful connections as scientists rushed to understand the pandemic. "Twitter was a really powerful way to do rapid science in some of the areas that we were working," says Carl Bergstrom, an evolutionary biologist at the University of Washington in Seattle. For instance, one of his most important early collaborators in trying to understand and model COVID-19 through Twitter was a hockey statistician, Bergstrom says.

At the same time, prominent COVID-19 researchers experienced insults, abuse and sometimes, as a 2021 *Nature* survey showed, death threats — often through Twitter (see *Nature* **598**, 250–253; 2021). Meanwhile, some researchers on the site oversimplified information, posted alarmist analyses or shared outright disinformation, Bergstrom adds. And despite Twitter's self-styled reputation as a public town square — where everyone gathers to see the same messages — in practice, the pandemic showed how users segregate to follow mostly those with similar views, argues information scientist Oliver Johnson at the University of Bristol, UK. For instance, those who believed that COVID-19 was a fiction would tend to follow others who agreed, he says, whereas others who argued that the way to deal with the pandemic was to lock down for a 'zero COVID' approach were in their own bubble.



Information scientist Oliver Johnson, who gained a Twitter following of more than 40,000 during the COVID-19 pandemic.Credit: Chrystal Cherniwchan

Bergstrom thinks the positives of Twitter outweighed the negatives. During the pandemic, it gave the public more transparency about the uncertain process of science progressing in real time, he says. And if some audiences wanted to leap on to messages of scientific certainty where there was none, that wasn't Twitter's fault, he adds.

"I don't think we've done a good job of talking in school science classes about the process of doing science, and explaining to people how the social process of science operates," he says. "When you actually see science in the making, it looks very, very different."

Days after Bergstrom spoke to *Nature,* however, he locked his own account after Musk's mocking tweet about Fauci. "You can't have meaningful and productive scientific collaboration on a platform run by [a] right-wing troll

who denies science when its results are inconvenient to him and just simply to hear his audience cheer," he wrote on Mastodon.



Evolutionary biologist Carl Bergstrom, who left Twitter in December 2022.Credit: Kris Tsujikawa

Public square, private land

Besides Musk's personal views, his changes to Twitter have worried plenty of scientists — particularly because he fired many people who work on content moderation. Scientists have noted, in particular, a Twitter announcement on 23 November that it would stop enforcing its COVID-19 misinformation policy. And there have since been reports that hate speech on the platform is increasing, including in areas such as climate science.

"We've been having conversations about if Twitter is now a safe place for our organization to exist, because of the way it's changing," says Bonner. "At this moment in time, I don't know."

Information scientist Stefanie Haustein at the University of Ottawa in Canada, who has studied the impact of Twitter on scientific communication, says the changes show why it's concerning that scientists embraced a private, for-profit firm's platform to communicate on. "We're in the hands of actors whose main interest is not the greater good for scholarly communication," she says.

Researchers leaving the platform will probably try to find a similar social-media replacement, says Rechavi. "I imagine that if Twitter stops being the place for scientists to be, then it'll be replaced by something else," he says. "I just can't imagine going back to being disconnected from the rest of the science world."

But Bonner says she doesn't think there's yet a space similar to Twitter. Dynamics on Instagram, where #BlackinCancer has a foothold, are drastically different, with less conversation and less reading of posts. And on Mastodon, the open-source alternative to Twitter that Bergstrom and Knight joined, users can post longer messages, but the dynamics of the platform deliberately make it harder to discover or encounter messages from users one doesn't directly follow, making communities more siloed and fragmented. (User numbers are still tiny compared with Twitter, estimated at some 2.5 million in early December.)

"A social network is always only successful if it's got enough people, and if

it’s got the right people,” says Haustein. “It requires millions of people to move from one place to the other.” Even if that happens, she says, you need to rebuild the same networks and structures that existed on Twitter — which is proving hard because of the way that control of Mastodon is distributed across servers, making it difficult for those who were on Twitter to reconnect.

Still, Quintana is hopeful: “Despite the fact that I’ve probably got ten times more followers on Twitter, the stuff that I posted is getting about the same amount of engagement on Mastodon,” he says.

For many, the tweet about Fauci was a final straw. Afterwards, a fresh wave of scientists decided to leave Twitter. But some are encouraging their colleagues to stick around. Rechavi emphasizes that Twitter has had a crucial role in research: “I hope it survives,” he says.

And, although the platform’s worst qualities are becoming more common, say researchers who spoke to *Nature* for this article, there is still a need for trained scientists to provide their expertise and point people to the best sources of evidence-based information. In reply to Bergstrom’s farewell, Trish Greenhalgh, a health scientist at the University of Oxford, UK, argued that people like him are still needed, and that she feels duty-bound to carry on: “We can and must stick around and post sensible scientific tweets. I’m staying.”

Nature **613**, 19-21 (2023)

doi: https://doi.org/10.1038/d41586-022-04506-6

References

1. Mongeon, P., Bowman, T. D. & Costas, R. Preprint at arXiv https://arxiv.org/abs/2208.11065 (2022).

2. Costas, R. *et al. Quant. Sci. Stud.* **1**, 771–791 (2020).

3. Fang, Z., Costas, R., Tian, W., Wang, X. & Wouters, P. *Scientometrics* **124**, 2519–2549 (2020).

4. Waltman, L. *et al*. *Scholarly Communication in Times of Crisis: The Response of the Scholarly Communication System to the COVID-19 Pandemic* (Research on Research Institute, 2021).

5. Fang, Z., Costas, R., Tian, W., Wang, X. & Wouters, P. *J. Assoc. Inf. Sci. Technol*. **72**, 918–932 (2021).

6. Vosoughi, S., Roy, D. & Aral, S. *Science* **359**, 1146–1151 (2018).

7. Ribeiro, M. H., Calais, P. H., Santos, Y. A., Almeida, V. A. F. & Meira, W. Jr *Proc. Int. AAAI Conf. Web Soc. Media* **12**, 676–679 (2018).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-022-04506-6

Opinion

- **Alfred Russel Wallace’s first expedition ended in flames** [04 Jan 2023]
  • Born 200 years ago, the evolutionary biologist experienced many setbacks during his career — none more severe than when he headed home with his precious collections from Brazil.
- **Escaping Darwin’s shadow: how Alfred Russel Wallace inspires Indigenous researchers** [04 Jan 2023]
  • Wallace, who independently discovered the theory of evolution, relied on local knowledge to craft his seminal work on species ranges in the Amazon. Now, the region’s Indigenous scientists have taken charge of their research using this and other cross-cultural tools.

- COMMENT
- 04 January 2023

Alfred Russel Wallace’s first expedition ended in flames

Born 200 years ago, the evolutionary biologist experienced many setbacks during his career — none more severe than when he headed home with his precious collections from Brazil.

- Andrew Berry



Naturalist Alfred Russel Wallace went on an expedition to Amazonas state in Brazil in 1848–52.Credit: Mondadori Portfolio via Getty

Best known for formulating the theory of evolution by natural selection, independently of Charles Darwin, Alfred Russel Wallace is an appealing if enigmatic figure. The appeal stems in part from his underdog status: poor and self-educated, Wallace had none of Darwin's social and financial advantages. The enigma comes from his keen embrace of a range of eccentric non-scientific causes, including spiritualism, phrenology and anti-vaccination (for smallpox).

Scientists do not like their scientific heroes to bear the taint of irrational thinking. Wallace's enthusiasms have therefore contributed to him becoming marginalized in the history of evolutionary thought. Most people know about Darwin and the HMS *Beagle*. But what about Wallace and the *Helen*?

The *Helen* story is worth revisiting because it shows Wallace at his resolute best. Despite numerous disastrous career setbacks — of which the *Helen* episode was the most severe — he persevered and eventually succeeded as a scientist.

More than 150 years after Wallace's experience on the *Helen*, doing science continues to be hard and can be disappointing. Wallace's misadventure provides both perspective and an object lesson in how to navigate setbacks. His response to problems showcases his most inspiring traits: his commitment to science, his almost superhuman resilience and his refusal to mire himself in self-pity.

Tropical explorations

In his first job as a land surveyor, Wallace developed an interest in the plants he encountered as he tramped across the countryside. Then, in 1844 at the age of 21, he met Henry Walter Bates, who would later discover 'Batesian mimicry' (whereby members of a palatable prey species gain protection by mimicking an unpalatable one).

Bates, two years Wallace's junior, had a fixation with beetles, and he catalysed Wallace's transformation from hobbyist naturalist to serious collector. Wallace's new-found focus on beetles transcended mere entomological stamp-collecting; he developed an interest in some of the great scientific questions of the time. He was particularly inspired by the anonymously published *Vestiges of the Natural History of Creation* (1844) by Robert Chambers, which put forward a vision of a transmutational process, with life progressing from simple to complex.

Without money or connections, Wallace and Bates aspired to careers in science at a time when the field was the preserve of the moneyed elite. They would have to fund their scientific explorations by collecting and selling specimens. After a hasty choice of destination — tropical South America — and a crash course in collecting methods, Wallace, aged 25, and Bates, aged 23, arrived in Belém, Brazil, in May 1848 (see 'Doggedly determined').

Doggedly determined

Alfred Russel Wallace tends to be unjustly relegated to a footnote in the Charles Darwin story. He was, in fact, a pioneering biologist who refused to let disadvantage or disaster prevent him from pursuing his scientific dreams.

January 1823: Alfred Russel Wallace is born in Usk in Wales.

May 1848: Wallace and Henry Walter Bates arrive in Belém, Brazil.

July 1852: Wallace boards the *Helen*, which catches fire three weeks later while at sea.

October 1852: Wallace reaches Deal, England, aboard the *Jordeson*.

March 1854: Wallace leaves Southampton for southeast Asia.

September 1855: Wallace's first evolutionary paper describing his 'Sarawak Law' is published.

May 1856: Citing the Sarawak Law paper, geologist Charles Lyell alerts Darwin to the possibility that Wallace is developing ideas similar to Darwin's.

February 1858: Wallace sends his paper on natural selection to Darwin from Ternate in the Maluku islands (Moluccas), Indonesia.

July 1858: The joint Darwin–Wallace paper is presented at the Linnean Society in London.

November 1859: Darwin's *On the Origin of Species* is published.

March 1862: Wallace returns from southeast Asia.

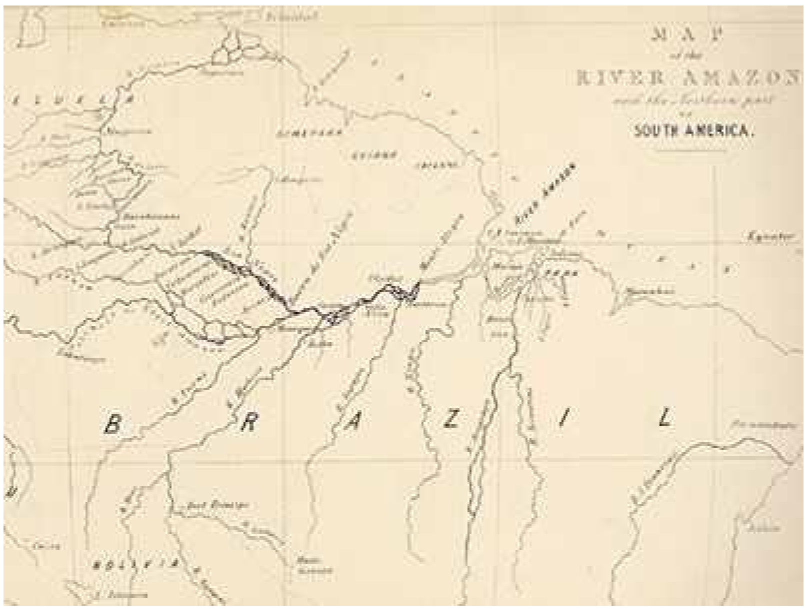
November 1913: Wallace dies in Broadstone, England.

The two split up early on, with Wallace concentrating on the Amazon River's northern tributary, the Rio Negro, and Bates on the southern fork, the Solimões.

Collecting was challenging. The Amazon's ubiquitous ants often deprived

science of hard-won specimens. Crucial collecting materials also disappeared: Wallace once recovered from a bout of fever to discover that local people had drunk the cachaça (a Brazilian rum) he'd been using to pickle specimens. Transport was a constant headache, with travel upstream past rapids requiring unwieldy portages of canoes and cargo. And thanks to his collecting, the cargo became ever more voluminous and unwieldy.

Wallace and Bates sporadically sent back shipments of material to their agent in London, Samuel Stevens, who publicized their adventures in scientific journals and sold their specimens, taking a 20% commission.



Escaping Darwin's shadow: how Alfred Russel Wallace inspires Indigenous researchers

Wallace's journeys on the Rio Negro and its tributaries took him into areas that had not yet been visited by Europeans. He saw (and collected) an extraordinary array of species, many of them new to science. He had a chance to observe and collect artefacts from several Indigenous groups with little or no previous contact with Europeans. As he travelled, Wallace capitalized on his surveying skills to map the terrain. But the remoteness took its toll. He made an "inward vow never to travel again in such wild, unpeopled districts without some civilised companion or attendant"[1].

Wallace was frequently ill, on one occasion nearly lethally so. His younger

brother came out to join him as an assistant in 1849 but died of yellow fever two years later in Belém, on his way back to England. Wallace learnt that his brother was sick but had to wait many anxious months before news of his death made it upriver.

In 1852, after four years of exploring and collecting, it was time for Wallace himself to head home. He envisaged a triumphant return. He would complement his collections of preserved organisms with a menagerie of living ones. Mr Wallace's biological wonders would surely be the toast of scientific London.

On 12 July in Belém, Wallace boarded the *Helen*, a freighter ship bound for London. The trip across the continent to Belém had not gone smoothly. The authorities in Manaus, Brazil, had had to be persuaded to release some of his earlier shipments meant for London, which they had impounded, making the final haul aboard the *Helen* even larger. But now all that remained was the long voyage back across the Atlantic. Wallace, who shared Captain Turner's cabin, was the only passenger.

Disaster strikes

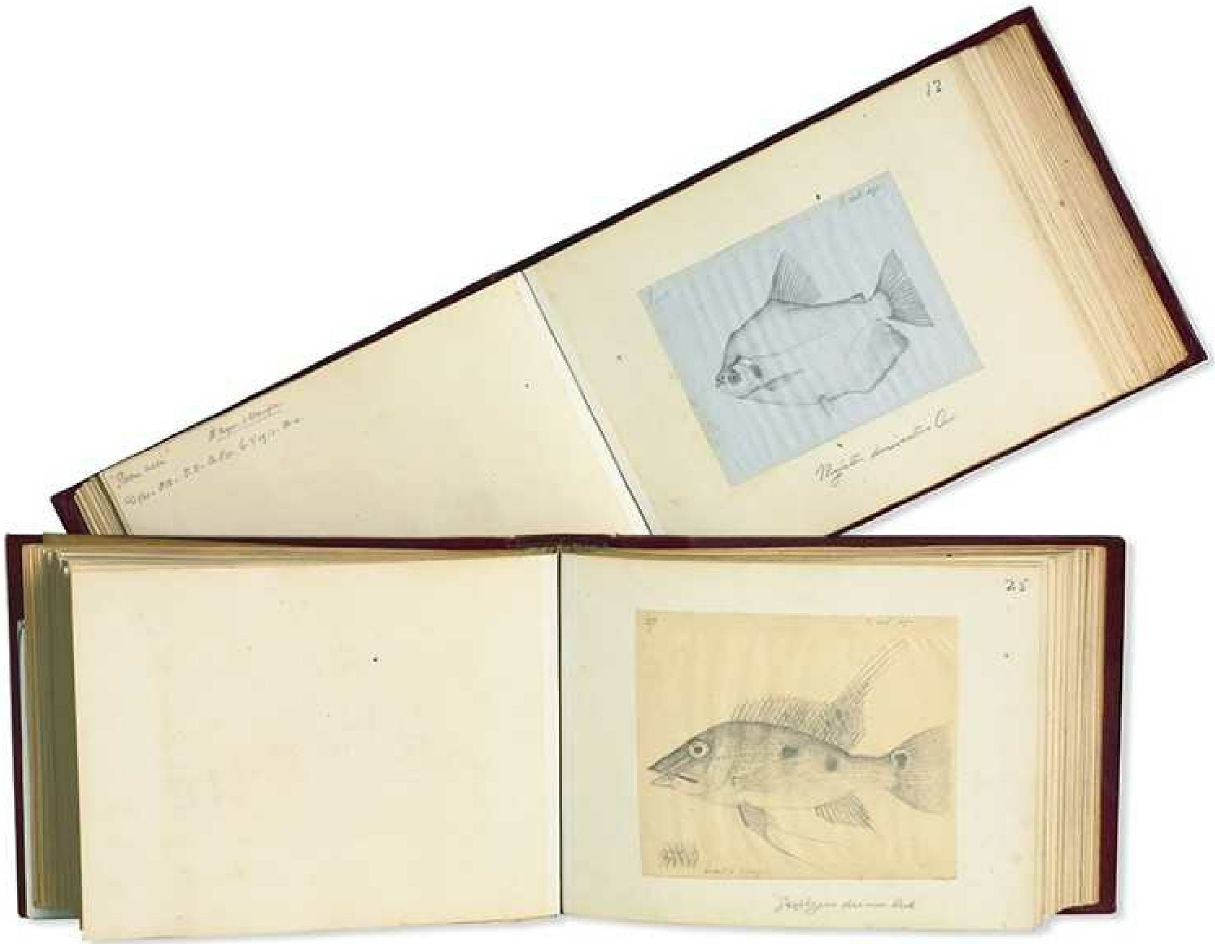
Three weeks into the voyage, Captain Turner interrupted Wallace's morning routine to tell him that the ship was on fire.

Friction caused by the rocking of the ship had ignited poorly stowed cargo. Attempts to intervene were counterproductive — removing the hold covers merely oxygenated the fire — and soon the ship became what Wallace later called "a most magnificent conflagration"[1].

Captain Turner gave the order to abandon ship, and the scramble to prepare two small wooden boats began. Having been stored on deck in the tropical sunshine, both boats leaked badly. The cook had to find corks to plug their hulls.

Before he left the ship, Wallace "went down into the cabin, now suffocatingly hot and full of smoke, to see what was worth saving"[1]. He retrieved his "watch and a small tin box containing some shirts and a couple

of old note-books, with some drawings of plants and animals, and scrambled up with them on deck"[1]. He tried to lower himself on a rope into one of the small boats, but fever-weakened, he ended up sliding down the rope, stripping the skin off his hands.



Some of Alfred Russel Wallace's sketches were salvaged from the fire aboard the *Helen* on his return journey from South America in 1852.Credit: The Natural History Museum/Alamy

With fine weather, the best hope of rescue lay in other ships seeing the fire. The two boats duly circled the burning wreck for the next 24 hours, meaning that Wallace got to witness every moment of the tragedy. The animals he had brought with him on the long river journey across the continent, now free from their cages, sought refuge on the one part of the ship still untouched by the flames, the bowsprit. Wallace watched as the monkeys, parrots and more

— his pets as well as his best hope of impressing London's scientific elite — were incinerated.

The hoped-for rescue did not immediately materialize, and Captain Turner turned the two open boats towards Bermuda, 1,100 kilometres away to the northwest.

As the days ticked by, the situation became increasingly desperate. Water ran low and the tropical sun left Wallace's "hands and face very much blistered"[1]. Wallace nevertheless remained upbeat, later recalling that during one night, he "saw several meteors, and in fact could not be in a better position for observing them, than lying on [his] back in a small boat in the middle of the Atlantic"[1].

Finally, ten days into the ordeal, salvation appeared on the horizon in the form of the *Jordeson*, a creaking and already overladen cargo ship bound for London.

With the immediate crisis past, the magnitude of what had happened started to sink in. In a letter[2] written aboard the *Jordeson* to botanist Richard Spruce (see go.nature.com/3prhbdk), Wallace tallied his catastrophic losses — "almost all the reward of my four years of privation & danger was lost" — and concluded with characteristic understatement, "I have some need of philosophic resignation to bear my fate with patience and equanimity."



Evolution’s red-hot radical

The *Jordeson* finally limped into Deal, England, on 1 October 1852. Wallace had been at sea for 80 days. His outward voyage with Bates had taken only 29 days.

Wallace added a PS to his letter to Spruce. First there was immediate exhilaration about the return — “Such a dinner! Oh! beef steaks & damson tart”. But then came thoughts about the future: “Fifty times since I left Pará [Belém] have I vowed if I once reached England never to trust myself more on the ocean.” Even then, he noted that “good resolutions soon fade”.

Stevens had thoughtfully taken out insurance. So Wallace had £200 (US$980 at the time) — a fraction of his collections’ actual value — to cover his costs for a year in London while he tried to salvage what he could from the disaster and make future plans.

He rushed out two books, one a travelogue, the other a more technical account of the palm trees of the Amazon. Neither did well — 250 copies remained unsold a decade later from the travel book’s print run of 750. But he was getting his name out there. Stevens, too, had a done a good job of publicizing Wallace’s discoveries while Wallace had been away.

Perhaps most crucially, the positive response of the UK Royal Geographical

Society to his mapping work of the Rio Negro yielded a free steamship ticket to Singapore.

In March 1854, less than 18 months since the *Jordeson*'s bedraggled arrival at Deal, Wallace departed from Southampton in England for what he would call the "central and controlling incident"[2] of his life.

Eight more years of perilous travel awaited. So, too, did the discoveries of what came to be known as Wallace's Line (a boundary between the Asian and Australasian biogeographic regions) and of the theory of evolution by natural selection[3,4].

The scientific acclaim that greeted Wallace's return from southeast Asia in 1862 was a just reward both for his contributions and for that phenomenal doggedness — his determination, despite everything, to be a scientist.

Nature **613**, 22-24 (2023)

doi: https://doi.org/10.1038/d41586-022-04507-5

References

1. Wallace, A. R. *Narrative of Travels on the Amazon and Rio Negro* (Reeve and Co., 1853).
2. Wallace, A. R. *My Life: A Record of Events and Opinions* Vol. 1 (Chapman and Hall, 1905).
3. Wallace, A. R. *Ann. Mag. Nat. Hist.* **16**, 184–196 (1855).
4. Darwin, C. R. & Wallace, A. R. *J. Proc. Linn. Soc.* **3**, 45–62 (1858).

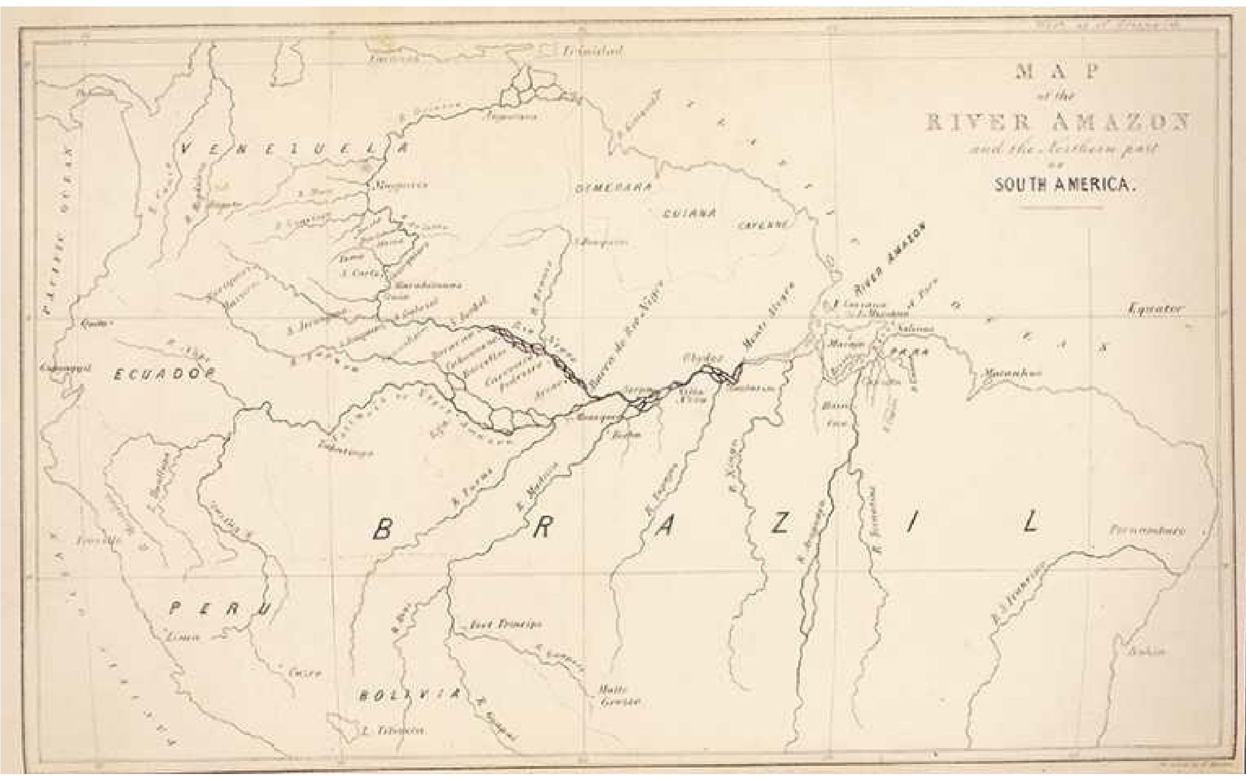
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-022-04507-5

- COMMENT
- 04 January 2023

Escaping Darwin’s shadow: how Alfred Russel Wallace inspires Indigenous researchers

Wallace, who independently discovered the theory of evolution, relied on local knowledge to craft his seminal work on species ranges in the Amazon. Now, the region’s Indigenous scientists have taken charge of their research using this and other cross-cultural tools.

- Camila C. Ribas



A map of the Amazon River and its tributaries, as published in Alfred Russel Wallace's 1853 book.Credit: Mary Evans/Natural History Museum

Dzoodzo Baniwa, a member of an Indigenous community in Brazil's Amazonas state, has been collecting data on the region's biodiversity for around 15 years. He lives in a remote village called Canadá on the Ayari River, a tributary of the Içana, which in turn feeds the Rio Negro, one of the main branches of the Amazon. The nearest city, São Gabriel da Cachoeira, is a three-day trip by motor boat.

Dzoodzo (who goes by his Indigenous name but is also known as Juvêncio Cardoso) takes inspiration for his work from many cross-cultural sources. A perhaps unexpected one is a 170-year-old book by the British naturalist Alfred Russel Wallace, who visited the Amazon and Negro rivers on his expeditions in 1848–52. *A Narrative of Travels on the Amazon and Rio Negro* gives detailed accounts of the wildlife and people Wallace encountered near Dzoodzo's home, including the Guianan cock-of-the-rock (*Rupicola rupicola*), a bright orange bird that Wallace describes as "magnificent … sitting amidst the gloom, shining out like a mass of brilliant flame"[1].

Dzoodzo's passion for local biodiversity is reflected in his work at Baniwa Eeno Hiepole School, an internationally praised education centre for Indigenous people. He dreams of one day turning it into a research institute and university that might increase scientific understanding of the region's species, including *R. rupicola*.



Alfred Russel Wallace’s first expedition ended in flames

Wallace, who was born 200 years ago, on 8 January 1823, is best known for spurring Charles Darwin into finally publishing *On the Origin of Species*, after Wallace sent Darwin his own independent discovery of evolution by natural selection in 1858. Most of Wallace’s subsequent work drew on observations from his 1854–62 expeditions in southeast Asia; his earlier work in Amazonia is much less well known.

Yet there are lessons from Wallace’s time in Brazil that are especially relevant for conservationists and other scientists today — notably, what can come from paying attention to what local people say about their own territory.

Barriers and boundaries

Wallace made two key contributions that still shape thinking about Amazonia, the world’s most biodiverse region, which covers parts of Bolivia, Brazil, Colombia, Ecuador, Peru, Venezuela, Guyana, Suriname and French Guiana.

On 14 December 1852, Wallace read out his manuscript ‘On the monkeys of the Amazon’ at a meeting of the Zoological Society of London. In this study,

which was later published[2], Wallace relays observations that form the basis of the most debated hypothesis for how Amazonian organisms diversified: the riverine barrier hypothesis.

His paper refers to the large Amazonian rivers as spatial boundaries to the ranges of several primate species. “I soon found that the Amazon, the Rio Negro and the Madeira formed the limits beyond which certain species never passed,” he writes. Since 1852, Wallace’s observations that large rivers could act as geographical barriers that shape the distribution of species have been corroborated, criticized and debated by many. The phenomenon he described clearly holds for some groups, such as monkeys and birds[3,4], but not for other groups, such as plants and insects[5].

Subsequent researchers have explored whether the distribution patterns of species, such as those observed by Wallace, indicate that the evolution of the Amazonian drainage system has itself driven the diversification of species[6]. Work in the past few years by geologists and biologists show that this drainage system, which includes some of the largest rivers in the world, is dynamic[7], and that its rearrangements lead to changes in the distribution ranges of species[8]. Current species ranges thus hold information about how the Amazonian landscape has changed over time.



The Guianan cock-of-the-rock (*Rupicola rupicol*a), which Wallace likened to a “brilliant flame”.Credit: Hein Nouwens/Getty

The second crucial observation made by Wallace, also in his 1852 paper, was that the composition of species varies in different regions. He describes how “several Guiana species come up to the Rio Negro and Amazon, but do not pass them; Brazilian species on the contrary reach but do not pass the Amazon to the north. Several Ecuador species from the east of the Andes reach down into the tongue of land between the Rio Negro and Upper Amazon, but pass neither of those rivers, and others from Peru are bounded on the north by the Upper Amazon, and on the east by the Madeira.” From these observations, he concluded that “there are four districts, the Guiana, the Ecuador, the Peru and the Brazil districts, whose boundaries on one side are determined by the rivers I have mentioned.”



Evolution’s red-hot radical

Even though Amazonia is presented as a single, large, green ellipse in most world maps, it is actually a heterogeneous place, with each region and habitat type holding a distinct set of species[9,10]. The four districts proposed by Wallace are bounded by the region’s largest rivers: the Amazon, Negro and Madeira. But further studies of species ranges since then have revealed more districts, now called areas of endemism, some of which are also bounded by these and other large Amazonian rivers, such as the Tapajós, Xingu and Tocantins[9,11].

This recognition of spatial heterogeneity in Amazonian species distributions — first accomplished by Wallace — is essential for today’s research, conservation and planning[10]. Each area of endemism includes species that occur only in that area. And different areas of endemism are affected differently by anthropogenic impacts, such as deforestation, fires and development[10]. More than half of Amazonia is now within federal or state reserves or Indigenous lands — territories that are recognized by current governments as belonging to Indigenous people. But nearly half of the region’s areas of endemism are located in the south of the region, close to the agricultural frontier, and the species they contain are severely threatened by habitat loss[10] (see also www.raisg.org/en).

Local knowledge

Although Wallace’s writings indicate that in many ways he admired most of the Indigenous people he met, especially those from the upper Rio Negro basin, he still viewed Indigenous people through the European colonial lens of his time. In *A Narrative of Travels on the Amazon and Rio Negro*[1], Wallace describes the Indigenous communities he encountered as “in an equally low state of civilization” — albeit seemingly “capable of being formed, by education and good government, into a peaceable and civilized community”.

Yet he did better than many of his contemporaries when it came to respecting local knowledge. In his 1852 paper, for example, Wallace notes that his fellow European naturalists often give vague information about the locality of their collected specimens, and fail to specify such localities in relation to river margins. By contrast, he writes, the “native hunters are perfectly acquainted” with the impact of rivers on the distribution of species, “and always cross over the river when they want to procure particular animals, which are found even on the river’s bank on one side, but never by any chance on the other.” Likewise, in his 1853 book[1], Wallace frequently corroborates his findings with information he has obtained from Indigenous people — for example, about the habitat preferences of umbrellabirds (*Cephalopterus ornatus*) or of “cow-fish” (manatees; *Trichechus inunguis*).

Considering the vastness and complexity of Amazonia, it is hard to see how Wallace could have gained the insights he did after working in the region for only four years, had he not paid close attention to local knowledge.



The other beetle-hunter

Amazonian Indigenous peoples have had to endure invasion of their lands, enslavement, violence from invaders and the imposition of other languages and cultures. Despite this, numerous Indigenous researchers wish to expand their knowledge about Amazonia by combining Indigenous and European world views. Meanwhile, a better understanding of how the Amazonian socio-ecological system is organized, and how it is being affected by climate change and local and regional impacts[12], hinges on the ability of researchers worldwide to learn from and to be led by Indigenous scientists.

The 98 Indigenous lands in the Rio Negro basin cover more than 33 million hectares (see go.nature.com/3wkkftu). If the hopes of Dzoodzo and others to build a research institute and university for the region are met, school students will no longer have to leave their homeland to pursue higher education. The community would have a way to document its own knowledge and that of its ancestors in a more systematic way. And the legitimization of Indigenous people's research efforts in the legal and academic frameworks recognized by non-Indigenous scientists — such as through the awarding of degrees — would make it easier for Indigenous researchers to partner with other organizations, both nationally and internationally.

Indigenous people in the Rio Negro basin today are no longer objects of observation — they have taken charge of their own research using tools from different cultures. Indeed, Dzoodzo is turning to Wallace's writings, in part, to learn more about how his own ancestors lived.

Perhaps the thread between Wallace and Dzoodzo, spanning so many years and such disparate cultures, could seed new kinds of partnership in which learning is reciprocal and for the benefit of all.

Nature **613**, 24-26 (2023)

doi: https://doi.org/10.1038/d41586-022-04508-4

References

1. Wallace, A. R. *A Narrative of Travels on the Amazon and Rio Negro, with an Account of the Native Tribes, and Observations on the Climate, Geology, and Natural History of the Amazon Valley* (Reeve and Co., 1853).

2. Wallace, A. R. *Ann. Mag. Nat. Hist.* **14**, 451–454 (1854).

3. Boubli, J. P. *et al. Mol. Phylogenet. Evol.* **82**, 400–412 (2015).

4. Ribas, C. C., Aleixo, A., Nogueira, A. C. R., Miyaki, C. Y. & Cracraft, J. *Proc. R. Soc. B* **279**, 681–689 (2012).

5. Dambros, C. *et al. Biodivers. Conserv.* **29**, 3609–3634 (2020).

6. Cracraft, J. *et al.* in *Neotropical Diversification: Patterns and Processes* (eds Rull, V. & Carnaval, A. C.) 225–244 (Springer, 2020).

7. Bicudo, T. C., Sacek, V., Paes de Almeida, R., Bates, J. M. & Ribas, C. C. *Sci. Rep.* **9**, 16879 (2019).

8. Musher, L. J. *et al. Sci. Adv.* **8**, eabn1099 (2022).

9. Guayasamin, J. M. *et al.* in *Amazon Assessment Report 2021* (eds

Nobre, C. *et al.*) Ch. 2 (Science Panel for the Amazon, 2021).

10. Ribas, C. C., Fritz, S. C. & Baker, P. A. *J. Biogeogr*. **49**, 1839–1847 (2022).

11. Cracraft, J. *Ornithol. Monogr*. **36**, 49–84 (1985).

12. Cabalzar, A. *Ciclos Anuais no Rio Tiquié* (Federação das Organizações Indígenas do Rio Negro & Instituto Socioambiental, 2016; in Portuguese).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-022-04508-4

Work

- **Classroom assistance: the scientists turning the tools of their trade to education** [03 Jan 2023]
  • A small but growing number of scientific faculty positions are focusing on the science of teaching.
- **Me and my rhino: a relationship five million years in the making** [03 Jan 2023]
  • A construction site in Tennessee that was once an ancient watering hole yields fossils that give field technician Laura Emmert clues about the rhinoceros, mastodon and other wildlife that congregated there.

- CAREER FEATURE
- 03 January 2023

Classroom assistance: the scientists turning the tools of their trade to education

A small but growing number of scientific faculty positions are focusing on the science of teaching.

- Amber Dance

Find a new job



Education-focused researchers explore the most effective ways of teaching science.Credit: Getty

As a doctoral student at the University of Washington in Seattle, biologist Michelle Smith spent a lot of time sorting fruit flies under the microscope. But she often found her mind wandering to her teaching activities: assisting with undergraduate laboratory and writing courses, and instructing schoolchildren in physics and biology.

“I felt this real draw towards teaching, but I also liked research,” she recalls. She felt lost, until 2007, when she took up a postdoctoral position in education research with Carl Wieman, a Nobel-prizewinning physicist with a deep interest in science education, then at the University of Colorado Boulder. She’d found her dream position: using her research skills to investigate how peer discussions help undergraduates to learn basic genetics concepts. She published her results in *Science*[1].

When she started her postdoc, Smith recalls, there weren’t many faculty positions available for education-focused researchers in biology departments. That’s changing, and she is now a tenured professor at Cornell University in Ithaca, New York, researching learning in ecology field courses, and senior associate dean for undergraduate education.



Stop describing academic teaching as a ‘load’

Tenure-track faculty members with expertise in education research are starting to pop up in science departments at leading research universities — although this remains the path less taken. Their presence, position and prestige also varies widely by nation, institution and field of study[2].

“There has been a growth in these types of position, both for tenure-track and long-term contract faculty members,” says Emily Miller, deputy vice-president for institutional policy at the Association of American Universities in Washington DC.

Precise data on the trend are scarce, but researchers from around the world point to the United States as a leader. The US National Science Foundation funds extensive research into science pedagogy, much of it through its Directorate for STEM Education, which doles out US$5 million to postdoctoral fellowships in science, technology, engineering and mathematics (STEM) education. There has also been a modest rise in a handful of other nations, including Canada and Australia, but they tend to have less money to support such studies, says Kimberly Tanner, a biologist at San Francisco State University in California, who focuses on biology-education research and co-edits the journal *CBE—Life Sciences Education*. The field has gained stature from the participation and support of big-name scientists such as Wieman and Tanner’s mentor Bruce Alberts. Alberts is a biochemist at the University of California, San Francisco, past president of the US National Academy of Sciences and author of the prominent textbook *The Molecular Biology of the Cell*.



Researcher Michelle Smith, who studies learning in ecology field courses at Cornell University in Ithaca, New York, guides students in an activity that models evolutionary principles.Credit: Cornell University

Faculty members in this speciality might largely teach, or they could maintain a strong research programme that focuses on education. “Within this field, you can think about the balance that you want, and you can have options,” says Smith.

Some researchers, such as Wieman, transitioned into education research after becoming established in their initial discipline. As the number of positions grows, early-career researchers will be able to enter education-focused positions directly. Faculty members who focus on education not only support their own students, but also help their colleagues to adopt the latest, science-based teaching methods, which in turn improves universities’ reputations with applicants.

Sideways steps

The discipline of physics has a long history of engaging with education research, says Wieman. His interest in education was sparked after starting his atomic-physics lab several decades ago, when he noticed that some graduate students who excelled at physics courses in the classroom were not as successful when they embarked on research in the lab. They were book smart, but lacked the problem-solving skills to debug scientific equipment or interpret experimental results. “There was just this fundamental puzzle,” he says. “Why were these things so disconnected?” So, for the next 15 years or so, he led parallel research programmes in atomic physics and physics education.



Better teachers are needed to improve science education

The atomic-physics programme earned him a Nobel Prize in Physics in 2001, and that also gave him a “bigger soapbox to stand on” to advocate for improvements in education. In 2007, he moved to the University of British Columbia in Vancouver, Canada, where he set up a Can$11-million (US$8-million) initiative to transform undergraduate science education with evidence-based methods, while managing a similar initiative in Colorado.

Wieman says that, rather than relying solely on lectures, teachers should

facilitate “learning to think in a different way, and there’s real expertise in how to guide people to do that”. Many education experts promote active-learning techniques, such as getting students to work together to solve problems. More than 100 courses at the University of British Columbia were altered as a result of Wieman’s initiative[3].

Wieman now has a joint appointment in the physics department and at the Graduate School of Education at Stanford University in California. Today, his research focuses exclusively on education, and specifically on undergraduate education of ‘technical expertise’ in physics, engineering and medicine.

Biology is another field with a history of education research, but it has taken time for many departments to prioritize educational innovation. In 1999, when evolutionary ecologist Raoul Mulder was interviewed for his first faculty job, a standard teaching and research position at the University of Melbourne in Australia, he recalls being asked only one question about lecturing.

As Mulder’s studies of bird ecology progressed, he, too, became intrigued by educational methods. “I think teaching is an absolutely natural outlet for this curiosity we all have as researchers,” says Mulder.

He began to turn the tools of science on education. He describes his initial education research, which began in the early 2000s, as “kind of a side hustle, I suppose. I wasn’t sure whether it was an accepted part of my identity”.

One of these side interests was student assessment. He noticed that students often didn’t pick up their final papers after grading. His careful feedback came too late to help them to improve their grades. So he experimented with a system wherein students received feedback from each other ahead of final submissions. Not only were the final papers improved, but students often couldn’t tell the difference between comments from their peers and their teachers[4].

As these side projects continued, Mulder accumulated education-focused grants adding up to nearly Aus$1 million (US$675,000). Although his education studies account for only about 10% of his papers, they fuel more

than 20% of his citations and make up nearly half of his top-ten most-impactful citations. “What was a side hustle has actually become more important than I thought,” Mulder says.



Improve undergraduate science education

It also marked him as someone interested in education studies, and in 2021 resulted in him taking on a leadership role at the Melbourne Centre for the Study of Higher Education, which focuses on evidence-based practices for higher education and professional development. Part of Mulder’s job includes handing out grants for others to do education research.

Those interested in the science of science education can start with questions inspired by their own classrooms, Mulder says, then dig into the “rich and growing” literature. He also suggests striking up conversations with other faculty members, who might be open to collaborating on studies of their own teaching practices.

But young researchers on a standard faculty track would be wise to speak with their department chair before focusing too much attention on education, says Neil Haave, a biologist at the University of Alberta in Camrose, Canada. “If they don’t see scholarship of teaching and learning as being research, that might not get you to full professor,” he warns.

Education-focused from the start

Another option is to enter an education-focused position directly as a new faculty member, as did Natasha Holmes, a physics-education researcher at Cornell who studies the efficacy of laboratory courses. “There are more and more graduate students coming up the pipeline now,” says Holmes, another Wieman protégé.

During her graduate studies at the University of British Columbia, Holmes assisted a postdoctoral fellow with an education research project, and was inspired to switch her PhD thesis to physics-education research. “It just tapped into my passions and curiosities,” Holmes recalls.



Researcher Natasha Holmes studies the efficacy of physics laboratory courses.Credit: Serge Petchenyi/Cornell University Center for Teaching Innovation

She focused on what she calls “a really extreme form of active learning”, when students invent solutions to problems before being taught the standard methods. It results in “much richer and deeper understanding”, Holmes says.

Tanner’s path, by contrast, started in the biology lab. She earned a PhD in neuroscience at the University of California, San Francisco, in 1997. She then did a postdoc in education research, studying partnerships between scientists and school classrooms. Tanner, who started her professor post at San Francisco State University in 2004, is particularly interested in how teachers and scientists can work together to make classroom biology more like actual biology research.

As her career has advanced, so has the community of researchers who share her passion. At a meeting of the Society for the Advancement of Biology Education Research last July, she saw her former postdocs and graduate students who now have teams of their own.

There are many journals, professional organizations and conferences at which people can start learning about education research (see ‘Resources for science-education researchers’). Social media is beneficial, too, says Manuel João Costa, deputy rector for student affairs and innovations in teaching and learning at the University of Minho in Braga, Portugal. He advises following the social-media accounts of teaching and learning centres, which often post content of interest.

Resources for science-education researchers

Scientists interested in pivoting to the science-education field can get started by reading the literature, attending conferences and joining relevant organizations.

Journals

• *Physical Review Physics Education Research*

• *CBE—Life Sciences Education*

• *Journal of Geoscience Education*

Professional organizations

• The International Society of the Learning Sciences

• Society for the Advancement of Biology Education Research

• American Association of Physics Teachers

• American Society for Engineering Education

• European Association for Research on Learning and Instruction

Conferences

Physics Education Research Conference in Sacramento, California, 19–20 July 2023

Gordon Research Conferences are small, international meetings at which scientists can discuss research that has not yet been published. Upcoming US events include:

• Undergraduate Biology Education Research in Lewiston, Maine, 25–30 June 2023

• Chemistry Education Research and Practice in Lewiston, Maine, 9–14 July 2023

• Visualization in Science and Education in Lewiston, Maine, 16–21 July 2023

Report

• *Reaching Students: What Research Says About Effective Instruction in Undergraduate Science and Engineering* (National Academies Press, 2015)

“It’s important to get some professional qualification in this space,” advises

Susan Rowland, deputy executive dean for the faculty of science at the University of Queensland in Brisbane, Australia. She gained a graduate certificate in teaching in 2007 as she pivoted towards an education-focused career, and says it helped her to become the first teaching-focused faculty member in the biochemistry department, in 2009. Teaching-specific training also schools researchers in the vocabulary that educators and education scholars use when they communicate and publish their work.

Being the first teaching-focused researcher in her department also meant that Rowland had to work out how to angle for tenure. To prove her worth, she says, it was crucial to operate in the ‘currency’ that other faculty members valued: papers, grants and visible public impact.

Eager educators should also be aware, Rowland adds, that by taking a teaching-focused position, they could be closing the door on a more conventional research-based career.

Seeking status

Perceptions of education and teaching vary, and Tanner has found that many academic deans in California exhibit some bias against education-focused researchers — although they appreciate the grant money these faculty members bring in[5]. Tanner recalls that numerous people told her that her career path was a waste of her neuroscience training. She says she was “pretty great” at neuroscience research, and that there is no reason to think that other education-minded scientists failed at the bench, either. She has obtained grants, awards and success by following her passions.

In fact, both students and faculty members benefit when education-focused academics are embedded in science departments. These specialists understand their science discipline better than do scholars who trained in education only, and they can serve as formal or informal resources for colleagues who want to improve in the classroom.

“We drive curriculum renewal and best practice in the classroom and online,” says Terry Mulhern, a biomedical-science educator at the University of Melbourne. “We innovate and improve, and in doing so we bring our

[department] colleagues with us.” For example, in 2019, he and a colleague revamped their university’s introductory biochemistry course using individualized feedback to students, lessons based on common misconceptions and several online elements. The latter proved useful when the institution had to pivot to full online learning in 2020 as a result of COVID-19 lockdowns.



Manuel João Costa teaches biochemistry to medicine students at the University of Minho in Braga, Portugal.Credit: Rui Oliveira

Changing colleagues’ attitudes towards education can take time, as Rowland knows well. When she was first hired to change the culture of teaching in the University of Queensland’s School of Chemistry and Molecular Biosciences, her first move was to choose the undesirable office next to the men’s toilets. Most of the department’s faculty were men, and she knew they’d all have to walk past her door eventually. That’s when she’d call out a cheery “Hi!”

After a while, colleagues started coming in, sitting in her comfortable chairs and talking about their students. Then they'd ask her questions — "You've published in *PNAS*? And in Nature journals? And you're teaching-focused?" Yes, yes and yes, she'd respond.

Eventually, Rowland says, she earned her peers' respect. Faculty members would come to her with teaching ideas, and she would help them to choose new approaches to try in their classrooms. "My presence legitimized the idea that they could talk about their courses as works in progress," she says. The biochemistry department now has 6 teaching-focused academics amid a faculty of 70.



[Ban the lecture](#)

At Cornell, Holmes says that she isn't formally expected to help her peers in the physics department to improve their teaching, but she serves as a resource just like any other faculty member with specific expertise. She knows cutting-edge teaching techniques she can implement and share with colleagues. For example, peer instruction is a technique in which lecture students answer a multiple-choice question, confer with their classmates, then answer it again, usually with improved understanding. The method often spreads between faculty members by word of mouth[6].

Holmes also brings a fresh perspective to education research, using

techniques such as machine learning. For example, a graduate student in her group applied theoretical statistical physics to model the messy data on student behaviours in physics lab courses. Using this technique, the researchers uncovered a gender divide in enquiry-based lab projects, with men more likely to use equipment and women more likely to analyse data[7].

All this effort in improving education eventually feeds the research enterprise, says Rui Oliveira, a biochemist and education researcher at the University of Minho. Well-taught undergraduates become capable graduate students and postdocs with critical-thinking skills — just what Wieman was after when he started investigating teaching 35 years ago.

“In the long run,” says Oliveira, “you will get better scientists to work with, and do better science.”

Nature **613**, 203-205 (2023)

doi: https://doi.org/10.1038/d41586-022-04509-3

References

1. Smith, M. K. *et al. Science* **323**, 122–124 (2009).
2. Bush, S. D. *et al. Proc. Natl Acad. Sci. USA* **110**, 7170–7175 (2013).
3. Wieman, C. *Improving How Universities Teach Science: Lessons from the Science Education Initiative* (Harvard Univ. Press, 2017).
4. Mulder, R., Baik, C., Naylor, R. & Pearce, J. *Assess. Eval. Higher Educ.* **39**, 657–677 (2013).
5. Bush, S. D., Stevens, M. T., Tanner, K. D. & Williams, K. S. *CBE—Life Sci. Educ.* **19**, ar34 (2020).
6. Dancy, M., Henderson, C. & Turpen, C. *Phys. Rev. Phys. Educ. Res.* **12**, 010110 (2016).
7. Quinn, K. N. *et al. Phys. Rev. Phys. Educ. Res.* **16**, 010129 (2020).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-022-04509-3

- WHERE I WORK
- 03 January 2023

Me and my rhino: a relationship five million years in the making

A construction site in Tennessee that was once an ancient watering hole yields fossils that give field technician Laura Emmert clues about the rhinoceros, mastodon and other wildlife that congregated there.

- Jack Leeming



Laura Emmert is a lab and field technician at the Gray Fossil Site and Museum in Gray, Tennessee. Credit: Mike Belleme

It’s going to take me three years to unearth this rhinoceros fossil, but that’s okay — he’s been waiting five million years for me to come along.

This particular fossil is of ‘Papaw’, part of a newly described species (*Teleoceras aepysoma*) from this site. He was very, very old when he died, around 30. He has arthritis in his knees. His teeth are broken. He’s got gouges in his front tusks, probably from fighting. This species is shorter than modern rhinos, but weighs as much or even more, around 2,000 kilograms.

After spending so much time with him, you start to get connected. You wonder about his life and times. Sometimes I chat to him as I dig: “You’ve got all your ribs together. You’re doing great.”

I’m working at the Gray Fossil Site in Gray, Tennessee, which was discovered in 2000 when a construction company tried to improve an intersection. They took off the top of a hill, found black clay and started hitting fossils. Since 2007, we’ve been operating a museum in addition to a dig site. Tourists can come and see the fossils right in front of them.

In Papaw’s time, this site was a watering hole, so many animals would congregate here. Besides the rhino, we’ve found a new species of mastodon (*Mammut* sp. nov.); a species of red panda (*Pristinailurus bristoli*) and many tapirs (*Tapirus polkensis*).

We avoid digs during the frost, so over winter I’ll be spending my time cleaning and categorizing the fossil fragments and putting the pieces back together to eventually reconstruct Papaw’s full skeleton.

I got a geology degree at East Tennessee State University and considered a PhD. But I decided to become a lab and field technician after I saw my advisers turn up at the dig site, see us working, and say “Oh, I wish I could dig, I’ve got to go back inside to write a grant.”

I decided that wasn’t for me. I’m happier here, chatting to Papaw and digging up bones.

Nature **613**, 208 (2023)

doi: https://doi.org/10.1038/d41586-022-04510-w

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-022-04510-w

Research

- ## Global estimates of excess deaths from COVID-19
  [14 Dec 2022]
  • Estimating the number of deaths attributable to COVID-19 around the world is a complex task — as highlighted by one attempt to measure global excess mortality in 2020 and 2021.
- ## The WHO estimates of excess mortality associated with the COVID-19 pandemic [14 Dec 2022]
  • Msemburi et al. describe how the World Health Organization has estimated the excess mortality associated with the COVID-19 pandemic, by month and for 2020 and 2021, and analyse their estimates across the WHO member states, with 14.83 million global excess deaths estimated.
- ## Recurrent repeat expansions in human cancer genomes [14 Dec 2022]
  • An atlas explores the landscape of recurrent repeat expansions in human cancer genomes.
- ## Cortical–hippocampal coupling during manifold exploration in motor cortex [14 Dec 2022]
  • A new study explores the activity dialogue between hippocampus and cortex as neuronal representations of learning or adaptation are consolidated.
- ## Microglia regulate central nervous system myelin growth and integrity [14 Dec 2022]
  • Resident microglia in the central nervous system are identified as the specific macrophage population that regulates myelin growth and integrity.
- ## Papers and patents are becoming less disruptive over time [04 Jan 2023]
  • A decline in disruptive science and technology over time is reported, representing a substantive shift in science and technology, which is attributed in part to the reliance on a narrower set of existing knowledge.
- ## Senescence atlas reveals an aged-like inflamed niche that blunts muscle regeneration [21 Dec 2022]
  • A lifetime cartography of in vivo senescent cells shows that they are heterogeneous. Senescent

cells create an aged-like inflamed niche that mirrors inflammation associated with ageing and arrests stem cell proliferation and tissue regeneration.

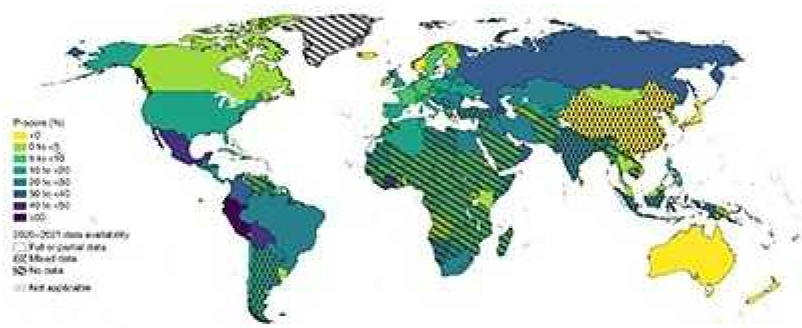
- NEWS AND VIEWS
- 14 December 2022

Global estimates of excess deaths from COVID-19

Estimating the number of deaths attributable to COVID-19 around the world is a complex task — as highlighted by one attempt to measure global excess mortality in 2020 and 2021.

- Enrique Acosta

Knowing how COVID-19 affects global mortality rates is crucial if we are to understand the factors that govern its spread and severity, and to be able to evaluate the effectiveness of government responses to the pandemic. In May, a team of researchers led by the World Health Organization (WHO) and the United Nations Department of Economic and Social Affairs published the first results from their attempt to estimate global, COVID-19-related death rates. Writing in *Nature*, Msemburi *et al*.[1] present these estimates in more detail.



[Read the paper: The WHO estimates of excess mortality associated with the COVID-19 pandemic]

Many deaths from COVID-19 went undetected in official reports from 2020 and 2021, because of limited testing capacity and misclassification of causes of death. This lack of data makes it challenging to quantify the mortality toll of short-term events, such as wars and natural disasters, as well as pandemics. For this reason, excess mortality — defined as the difference between all observed and expected deaths in a given period — is considered the gold-standard approach for estimating the mortality toll of short-term events[2,3]. But it is hard to find a universally effective way to measure excess mortality[4,5], because there are substantial variations in underlying mortality trends and data availability across populations.

Msemburi and colleagues set out to estimate excess deaths from COVID-19 for every country in the world. The authors report that there were between 13.2 million and 16.6 million more deaths than expected in 2020 and 2021. This death toll was between 2.4 and 3.1 times higher than the officially reported number of COVID-19-related deaths. Four out of five excess deaths occurred in middle-income countries (Fig. 1), with some of the worst affected in Latin America. In both years combined, the observed mortality in Peru was double the expected level, and it was between 41% and 51% higher than expected in Mexico, Bolivia and Ecuador. Low-income countries had fewer

deaths, mostly because they account for only 9% of the global population, and have younger populations, on average, than do higher-income countries. Compared with the published data, Msemburi *et al*. present more-detailed methods and more-refined estimates, by adjusting the way in which underlying mortality trends were forecast for several countries[6].

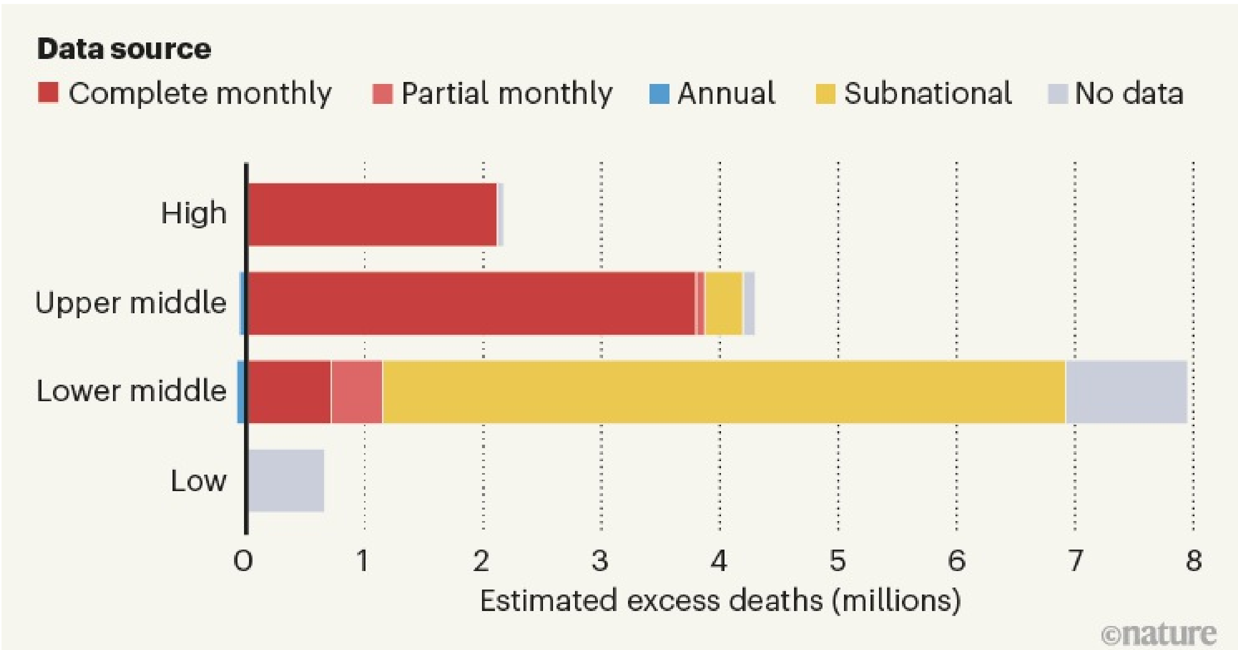
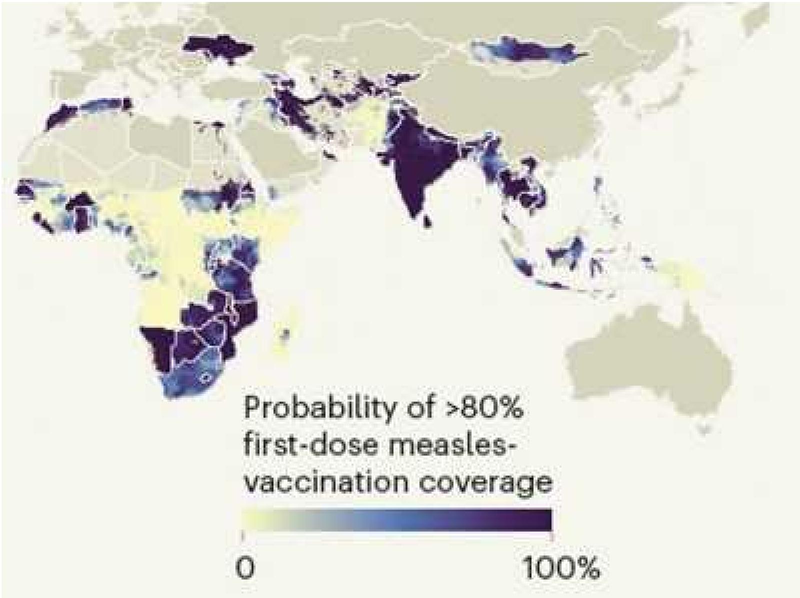


Figure 1 | Country income level and estimated excess deaths from COVID-19. Msemburi *et al*.[1] estimated the number of excess deaths that occurred in every country in the world in 2020 and 2021. Here, countries are grouped according to their income level (high, upper middle and so on). The authors made their estimates using monthly, country-level data on the total number of deaths where available. However, they had to make estimates of excess deaths from incomplete data for most countries. For some, data were available for only certain months or as annual data. More problematically, in many countries, data were available on only a subnational level, or not at all. Globally, half of the estimated excess deaths came from these countries, with this proportion varying hugely by the country's income level. The authors' estimates involved subtracting negative excess (where there were fewer than expected deaths) from positive excess, meaning that for some subgroups (such as annual upper middle and lower middle), the overall number of deaths is negative. (Figure generated using data from Supplementary Tables 5–13 of ref. 1.)

However, the authors' estimates must be interpreted with extreme caution. This is because only 37% of countries had complete data for the number of people who died from any cause in each month of 2020 and 2021 — an essential figure for accurately calculating excess mortality. There were no data at all for 43% of countries; 2% of countries had data for only some regions; 5% had only annual data; and 13% had incomplete monthly series. The availability of data is heavily correlated with income. For example, only 2% of European countries had no data for 2020, whereas for African countries this figure was 87%.

As a result, the authors had to make some problematic inferences. First, for countries that lacked data in some regions, they had to scale death counts from subnational to national levels. This assumes a constant share of deaths between regions before and during the pandemic. But the spread, timing and severity of COVID-19 within countries was far from uniform[7,8].



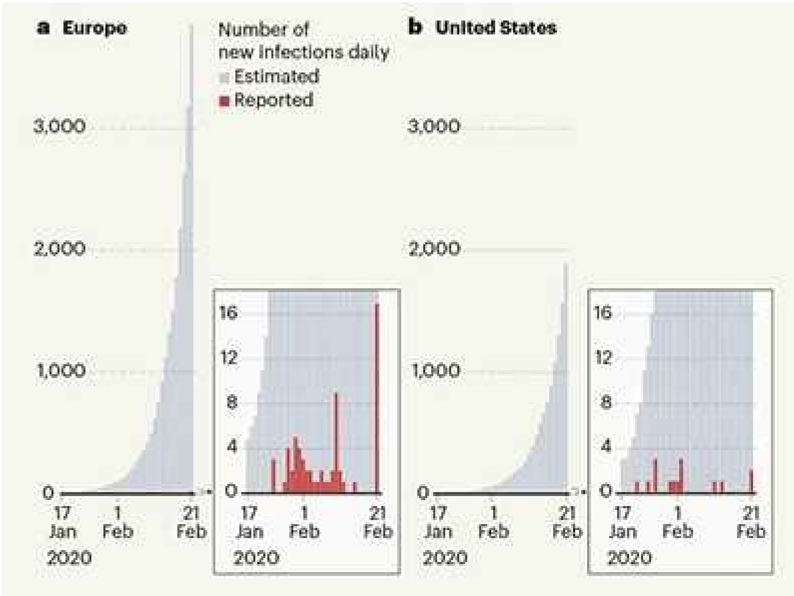
Precise mapping reveals gaps in global measles vaccination coverage

Second, the researchers had to infer the number of expected and excess deaths in countries with no mortality data (mostly low-income countries) by extrapolating patterns from those with more-complete data (mostly high-income countries that have robust health-care systems). The researchers adjusted the patterns seen across the latter group to make estimates for the

former using country-specific proxies for socio-economic conditions, the intensity of the pandemic, each country's susceptibility to COVID-19 and its capacity to respond to the crisis. As such, half of global excess deaths were estimated without data on mortality, or by using data from subnational regions (Fig. 1).

There are other factors to consider when interpreting Msemburi and colleagues' findings, including 'avoided' and 'displaced' mortality. The former refers to expected deaths that did not occur — for instance, influenza-related deaths that were avoided in 2020 and 2021, owing to changes in people's behaviour and social-isolation measures[9]. The latter refers to deaths of frail or sick people that were expected to occur during the observation period, but were brought forward by COVID-19, producing a temporary surplus in mortality followed by a deficit, and consequently ignored when calculating cumulative excess deaths.

Msemburi *et al*. did not adjust for these factors, because they aimed to identify all changes in mortality during the pandemic. Nevertheless, adjusting estimates for these numbers is essential if the goal is to understand COVID-19 fatality better. There are already models available that can adjust for influenza-related avoided mortality[10], and new approaches have been proposed for use in the context of the COVID-19 pandemic[11]. For displaced mortality, partial adjustments can be made by excluding mortality deficits when computing the cumulative number of excess deaths over the months of the observation period. (The adjustment is only partial because COVID-19 deaths that occur during periods of overall deficit remain uncounted.)



A reconstruction of early cryptic COVID spread

Although the inferences made by Msemburi and colleagues are not ideal, there is no obvious alternative. However speculative these estimates, most are surely closer to the truth than are officially reported numbers of deaths from COVID-19. To rely on confirmed deaths would imply that the pandemic spared low-income and lower-middle-income countries — vulnerable populations that have limited capacity for testing and response. This assumption is highly implausible, and even irresponsible.

Nonetheless, the complexity of the task is apparent from the fact that similar attempts to estimate the effect of the pandemic on global mortality have given different results. For the same period, the University of Washington's Institute for Health Metrics and Evaluation in Seattle[12] estimated 18.2 million excess deaths, and *The Economist* magazine estimated 16 million (see go.nature.com/3uykedp). This makes the WHO estimate the most conservative of the three. Compared with these other studies, the approach adopted by Msemburi *et al*. is simpler, and their estimation of uncertainties is more rigorous.

A sensible next step for Msemburi and colleagues — and one currently under way (see go.nature.com/3vhmybu) — is to include information on age in their data. The risk of COVID-19 death increases with age[13]. Because the

authors calculated excess mortality for entire populations, any difference between countries is affected by variations in their age compositions. There have been some attempts to rank countries' response to the pandemic (including one in the current study) by estimated overall death toll, but without data on age, any assessments of differences in pandemic severity or the effectiveness of responses will be biased.

Finally, the complexity of estimating the effect of the pandemic on global mortality underscores the urgent need to build robust, centralized systems that allow for real-time monitoring of global mortality. The construction of such systems will require considerable global efforts to strengthen civil registration and crucial statistics systems worldwide, especially in low- and middle-income countries. But, once built, they will serve as an essential early warning for future pandemics and health crises.

Nature **613**, 31-33 (2023)

doi: https://doi.org/10.1038/d41586-022-04138-w

References

1. Msemburi, W. *et al. Nature* **613**, 130–137 (2023).
2. Leon, D. A. *et al. Lancet* **395**, e81 (2020).
3. Helleringer, S. & Queiroz, B. L. *Int. J. Epidemiol.* **51**, 85–87 (2022).
4. Nepomuceno, M. R., Klimkin, I., Jdanov, D. A., Alustiza-Galarza, A. & Shkolnikov, V. M. *Popul. Dev. Rev.* **48**, 279–302 (2022).
5. Schöley, J. Preprint at medRxiv https://doi.org/10.1101/2021.06.04.21258353 (2021).
6. *Nature* **606**, 242–244 (2022).
7. Lima, E. E. C. *et al. Genus* **77**, 30 (2021).
8. Konstantinoudis, G. *et al. Nature Commun.* **13**, 482 (2022).

9. Simonsen, L. & Viboud, C. *eLife* **10**, e71974 (2021).
10. Thompson, W. W. *et al. Influenza Other Respir. Viruses* **3**, 37–49 (2009).
11. Shkolnikov, V. M. *et al. SSM Popul. Health* **18**, 101118 (2022).
12. Wang, H. *et al. Lancet* **399**, 1513–1536 (2022).
13. Sasson, I. *Demogr. Res.* **44**, 379–396 (2021).

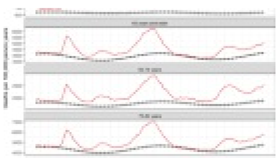
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-022-04138-w



Variation in excess all-cause mortality by age, sex, and province during the first wave of the COVID-19 pandemic in Italy

20 January 2022

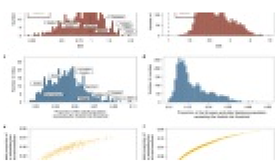
Nathaniel J. Henry, Ahmed Elagali, … Catrin E. Moore



Excess all-cause mortality in the USA and Europe during the COVID-19 pandemic, 2020 and 2021

03 November 2022

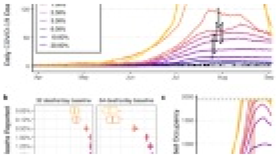
Lauren M. Rossen, Sarah K. Nørgaard, … Jens Nielsen



Individual and community-level risk for COVID-19 mortality in the United States

11 December 2020

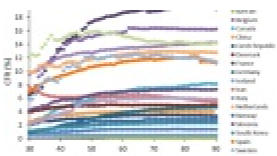
Jin Jin, Neha Agarwala, … Nilanjan Chatterjee



[Leveraging community mortality indicators to infer COVID-19 mortality and transmission dynamics in Damascus, Syria]

22 April 2021

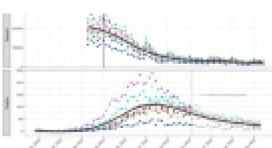
Oliver J. Watson, Mervat Alhaffar, … Patrick Walker



[Explaining among-country variation in COVID-19 case fatality rate]

03 November 2020

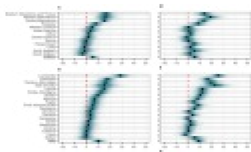
Gabriele Sorci, Bruno Faivre & Serge Morand



[Analysis of temporal trends in potential COVID-19 cases reported through NHS Pathways England]

29 March 2021

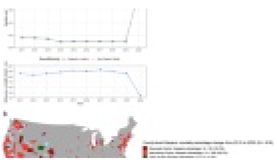
Quentin J. Leclerc, Emily S. Nightingale, … Thibaut Jombart



Regional excess mortality during the 2020 COVID-19 pandemic in five European countries

25 January 2022

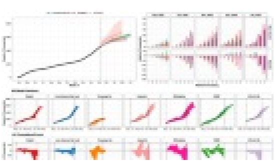
Garyfallos Konstantinoudis, Michela Cameletti, … Marta Blangiardo



Trends in national and county-level Hispanic mortality in the United States, 2011–2020

12 July 2022

Ashish Sarraju, Summer Ngo, … Fatima Rodriguez



Predictive performance of international COVID-19 mortality forecasting models

10 May 2021

Joseph Friedman, Patrick Liu, … Emmanuela Gakidou

- Analysis
- Open Access
- Published: 14 December 2022

The WHO estimates of excess mortality associated with the COVID-19 pandemic

- William Msemburi ,
- Ariel Karlinsky,
- Victoria Knutson,
- Serge Aleshin-Guendel,
- Somnath Chatterji &
- …
- Jon Wakefield

Show authors

Nature **volume 613**, pages 130–137 (2023)

- 71k Accesses
- 1 Citations
- 3067 Altmetric
- Metrics details

Abstract

The World Health Organization has a mandate to compile and disseminate statistics on mortality, and we have been tracking the progression of the COVID-19 pandemic since the beginning of 2020[1]. Reported statistics on

COVID-19 mortality are problematic for many countries owing to variations in testing access, differential diagnostic capacity and inconsistent certification of COVID-19 as cause of death. Beyond what is directly attributable to it, the pandemic has caused extensive collateral damage that has led to losses of lives and livelihoods. Here we report a comprehensive and consistent measurement of the impact of the COVID-19 pandemic by estimating excess deaths, by month, for 2020 and 2021. We predict the pandemic period all-cause deaths in locations lacking complete reported data using an overdispersed Poisson count framework that applies Bayesian inference techniques to quantify uncertainty. We estimate 14.83 million excess deaths globally, 2.74 times more deaths than the 5.42 million reported as due to COVID-19 for the period. There are wide variations in the excess death estimates across the six World Health Organization regions. We describe the data and methods used to generate these estimates and highlight the need for better reporting where gaps persist. We discuss various summary measures, and the hazards of ranking countries' epidemic responses.

Main

The COVID-19 pandemic caught the world unprepared, and it has exacted a toll many would have considered inconceivable in the modern era before its emergence. As of 31 December 2021, more than 287 million confirmed cases of COVID-19 across the world had been reported to the World Health Organization (WHO) including 5.4 million deaths (https://covid19.who.int/).

From the documenting of initial COVID-19 cases in Wuhan, China in December of 2019 through to the WHO declaring it a pandemic in March of 2020[1], accurately tracking COVID-19 and its impact has been riddled with challenges. An initial major challenge was developing the diagnostic tools to correctly identify the presence of the virus. A number of countries relied on pre-existing platforms to achieve this and they quickly adapted and scaled up the available technologies to allow for COVID-19 testing. However, many countries lacked such capacity. In addition, countries have differed in their application of standards for the certification of COVID-19 as the underlying cause-of-death[2]. This has caused both country-level and worldwide assessment of the spread and impact of the pandemic to be incomplete. An estimate of the excess mortality associated with the COVID-19 pandemic is

therefore a better measure of the overall impact of the crisis.

Excess mortality is defined as, “The difference in the total number of deaths in a crisis compared to those expected under normal conditions”[3]. Excess mortality accounts for both the total number of deaths directly attributed to the virus and those resulting from the indirect impact, such as disruption to essential health services or travel disruptions[4]. Excess mortality is a well established concept dating back centuries[5], and has been used extensively to estimate the toll of past health crises and pandemics such as the 1918 ‘Spanish Flu’[6]. The measure overcomes the variation among countries in reporting and testing and the misclassification of the cause of death on death certificates[7] and requires only information on the total number of deaths during the health crisis, and before, to establish the expected number of deaths.

Unfortunately, excess mortality cannot be directly estimated for all countries owing to many not having the requisite all-cause mortality (ACM) data. The WHO usually receives routine mortality data on an annual basis following the year of death or after an even longer lag. Civil registration and vital statistics (CRVS) systems differ across countries with varying completeness, timeliness and quality control measures for compiling unit record cause of death numbers into aggregates identified by cause, age, sex, place and period of death[8,9,10,11]. Moreover, differential reporting coverage, the absence of electronic surveillance systems in some locations and limited investments in CRVS systems has resulted in many nations lacking the structures necessary to provide good-quality routine data, even before the onset of the pandemic. Correspondingly, they lack the capacity and data required to monitor ACM during this unprecedented pandemic. This results in numerous countries being unable to contribute to the centralized systematic mortality surveillance that would be necessary for the WHO to measure global-, regional- and country-level excess mortality. Acknowledging these data gaps, a model-based framework, relying on ACM information from countries for which data exist and other relevant factors, has been developed by the WHO. The purpose of this framework is to estimate country, regional and global excess deaths from 1 January 2020 to 31 December 2021 on a monthly timescale.

Process, methods and data

Detailed descriptions of the process followed, data used and methods applied to generate the estimates of excess mortality within this paper are provided in the Methods and summarized here. The WHO, in collaboration with the United Nations Department of Economic and Social Affairs, assembled a technical advisory group to develop guidance on how to best estimate excess mortality in light of extensive data gaps. Preliminary estimates were generated according to the advice of this group and, following standard WHO procedures, a consultative process was initiated with member states who assessed the input data sources, methods and results, and provided feedback which was in turn used to update the estimates. Ideally, we would have ACM data for all countries and for all months. The reality is that such monthly national data are available for only 100 countries (52%), with other countries having annual data, subnational data or no data. For the latter three cases, we predict the monthly data within a Poisson count model framework, as detailed in the Methods.

The ACM data used for our modelling come from various sources including nominated country focal points and public-facing databases, and can be characterized according to three criteria: whether the data are nationally representative, what proportion of the two-year pandemic period they cover and the time periods to which the data are aggregated.

By region, the monthly data that are available consist of only 6 out of the 47 countries in the Africa region (13%), 23 out of the 35 countries from the region of the Americas (66%), 9 out of the 21 countries from the Eastern Mediterranean region (43%), 51 out of the 53 countries in the European region (96%), 2 out of 11 countries in the South East Asia region (18%) and 9 out of 27 countries in the Western Pacific region (33%). There are 10 countries with subnational or annual data; 4 are from the region of the Americas, 1 from the European region, 3 from the South East Asia region and 2 from the Western Pacific region.

We use a Bayesian Poisson framework to estimate both the expected deaths (for all countries and all months) in the absence of the pandemic, and the ACM for those countries with no such data during the pandemic. In addition to the data on ACM, we gathered information on specific variables with spatiotemporal variations considered to be associated with changes in excess

mortality over the course of the pandemic. These variables are chosen based on the strengths of the associations and availability across locations for the duration under study. We consider several that are assumed to change by month such as the COVID-19 death rate, the COVID-19 test positivity rate, aggregate containment measures (combining lockdown restrictions and closures) and average national temperature, together with others that are fixed over the period of study including a high-income country binary indicator, historic cardiovascular disease death rates and historic diabetes prevalence rates. A log-linear regression model on these variables, also within the Bayesian Poisson framework, is used to predict mortality levels in the locations without adequate reporting of mortality during the pandemic. For a handful of countries, instead of covariates, their subnational observed deaths are used to predict the national deaths using multinomial models that assume the relationships estimated between pre-pandemic subnational and national mortality persist into the pandemic. Finally, the reported or distributions of predicted deaths, conditional on data availability, together with the derived expected death distributions, are used to estimate monthly excess deaths in all locations for the years 2020 and 2021.

There are many different facets to the monthly time series of excess mortality that we may examine: (1) the raw excess counts, (2) the excess rate (say per 100,000 of population), (3) the P-score, which is the ratio of the excess to the expected and (4) the ratio of the excess to the reported COVID-19 deaths. Each of the four metrics are not known with certainty, even when we observe the total ACM over the complete pandemic, because the expected numbers are estimated. In general, between-region and between-country comparisons are difficult for many reasons, including the different age structures of the populations.

Raw excess counts directly show the basic human toll of the pandemic but are obviously critically dependent on the population size of the group (for example, country, region) over which we are calculating the excess. The excess rate adjusts for the population, to produce summaries that are more comparable across countries. It would be preferable to adjust for the age–sex population structure of each country, but unfortunately we do not have reliable excess counts by age and sex for all countries of the world.

Another method to compare countries' excess deaths is to normalize the excess estimates by the expected number of deaths for the analysed period, expressed as a percentage. This measure is know as a P-score[12]. For example, if 100 deaths were expected to occur and the actual number of deaths was 140, excess deaths would be 40 and the P-score would be 40%. The P-score implicitly considers both the population size and the age structure. Two countries may have identical population sizes, but very different routine mortality because of the age structure. For example, both Iran and Germany have a similar population size of about 83 million in 2019 but annual mortality that year was almost 2.5 times higher in Germany[13], mainly due to the German population being much older. For example, in Germany, 16% of the population are over the age of 65 whereas in Iran it is less than 4% (ref. [14]). Hence, the expected deaths are higher in Germany than in Iran. Continuing with this example, an excess deaths estimate of (say) 100,000 in both countries would rank them identically on a per capita basis, but the P-score in Iran would be higher than Germany's owing to it being a higher relative increase compared to the expected number of deaths.

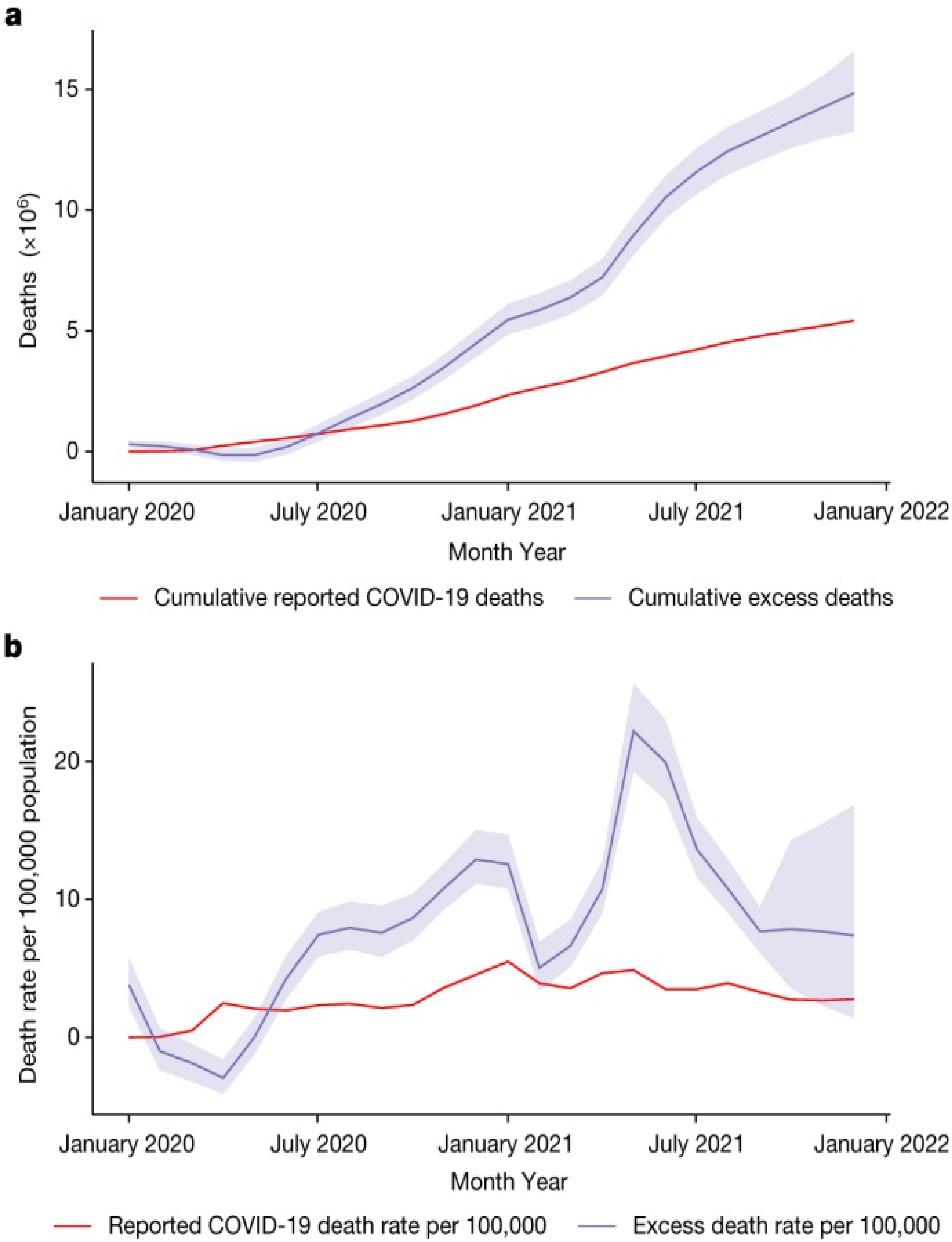
Global, regional and income group summary

Globally, for the period January 2020 to December 2021, we estimate 14.83 million excess deaths with an uncertainty interval (UI) of 13.23 million to 16.58 million, which is 2.74 (UI 2.44 to 3.06) times higher than the 5.42 million COVID-19 deaths reported to the WHO for this period. Throughout the paper, the reported UIs are 95% Bayesian credible intervals. We estimate 4.47 (UI 3.91 to 5.07) excess deaths in 2020 and 10.36 (UI 9.06 to 11.97) in 2021 globally. Turning to the P-scores, there were 7.97% (UI 6.96% to 9.03%) and 18.30% (UI 15.99% to 21.15%) increases in deaths globally in 2020 and 2021, respectively, compared to what we would have expected if the pandemic had not occurred.

The top panel of Fig. 1 shows the cumulative excess and reported COVID-19 deaths by month. We see the steepening of the curve in the middle of 2021. This sharp increase is almost entirely due to the estimate of the catastrophic wave that hit India at this time. In the bottom panel, the monthly excess death rate per 100,000 is plotted, and again, the peak towards the middle of 2021 is

evident.

Fig. 1: Global excess and reported COVID-19 deaths and death rates per 100,000 population.



a, Cumulative global excess death estimates and the cumulative reported

COVID-19 deaths by month from January 2020 to December 2021. **b**, Global excess death rates per 100,000 population and the reported COVID-19 death rates per 100,000 population, also by month, from January 2020 to December 2021. On both plots, the central lines of the excess mortality series show the mean estimates and the shaded regions indicate the 95% uncertainty intervals.

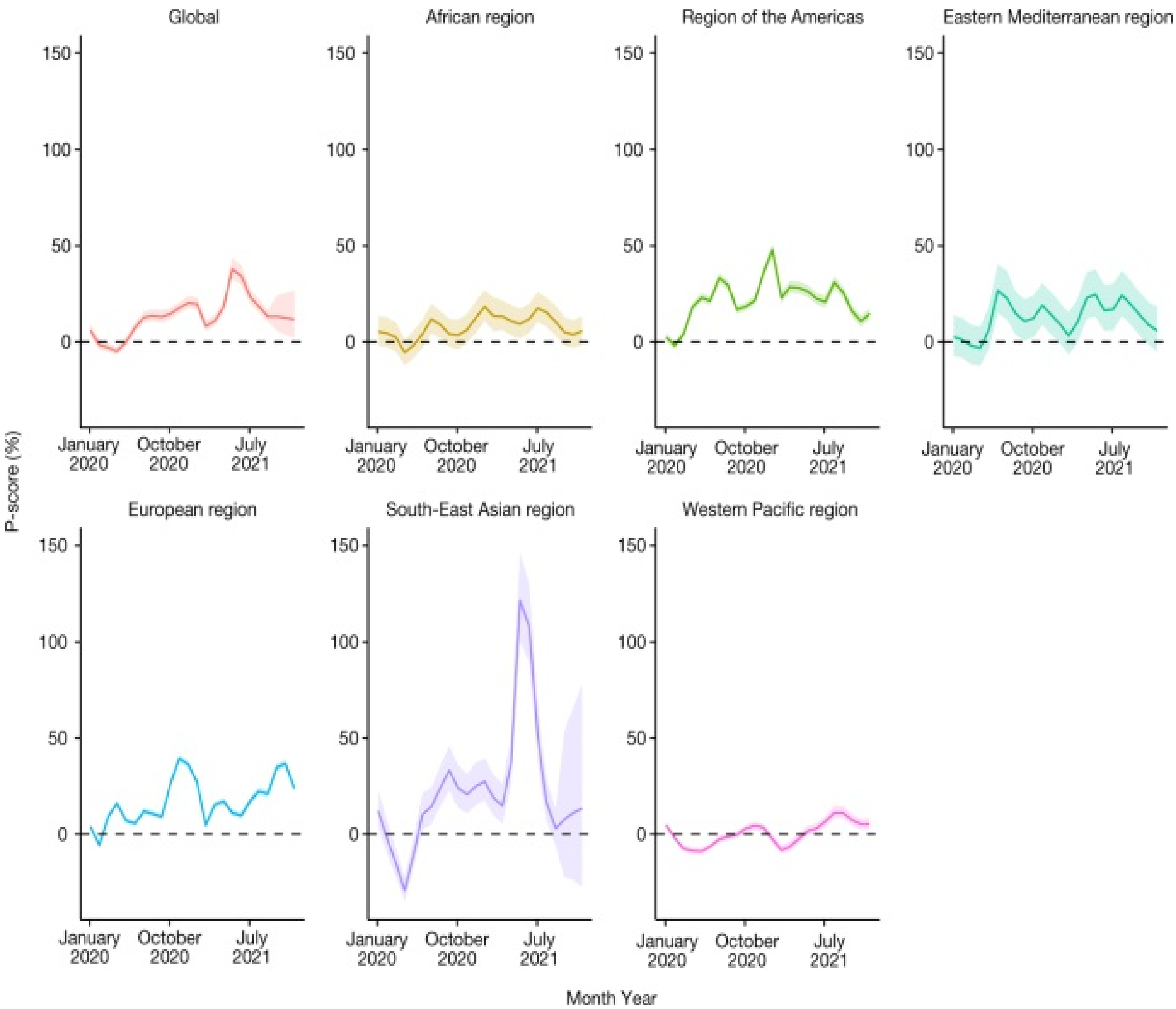
Source data

In the Extended Data Table 1 we provide summaries aggregated across all 194 WHO member states to give global estimates for the period January 2020 to December 2021. We also aggregate the member state estimates according to the six WHO regions and for the same period. The greatest contribution to the total is estimated for the SEAR region. The ratio of excess mortality to reported COVID-19 mortality is greatest in the AFR and SEAR regions, although we must emphasize that these are the regions with the greatest data paucity. As measured by the P-score, the worst affected regions were AMR (22%) and SEAR (22%), with EUR (17%) and EMR (12%) having intermediate values and AFR (8%) and WPR (0%) having the lowest values.

In the Extended Data Table 2 we also provide the corresponding summaries when the member state estimates are aggregated according to the four World Bank income groupings. There are wide variations across the respective economies but in general the estimates point to excess mortality being magnitudes higher than the reported COVID-19 mortality. Over 50% of the estimated excess occurs in lower-middle income economies. The low-income economies include much of sub-Saharan Africa, which reported relatively few COVID-19 deaths.

In Fig. 2 we plot P-scores by month globally and for each WHO region. The global P-score increases steadily to the end of 2020, and then drops before a sharp increase in the middle of 2021, followed by a steady decline. The striking peak is in SEAR for the middle of 2021 for which we estimate that more than twice as many deaths have occurred, relative to those expected for this period. AMR has a peak in January 2021, and EUR has a peak at the end of 2020, and then drops before steadily increasing in 2021.

Fig. 2: Global and WHO region P-scores (excess deaths relative to expected deaths).



Monthly estimates of P-scores, expressed as a percentage, aggregated globally and for the six WHO regions for the period January 2020 to December 2021. All plots show the mean estimates and the 95% uncertainty intervals.

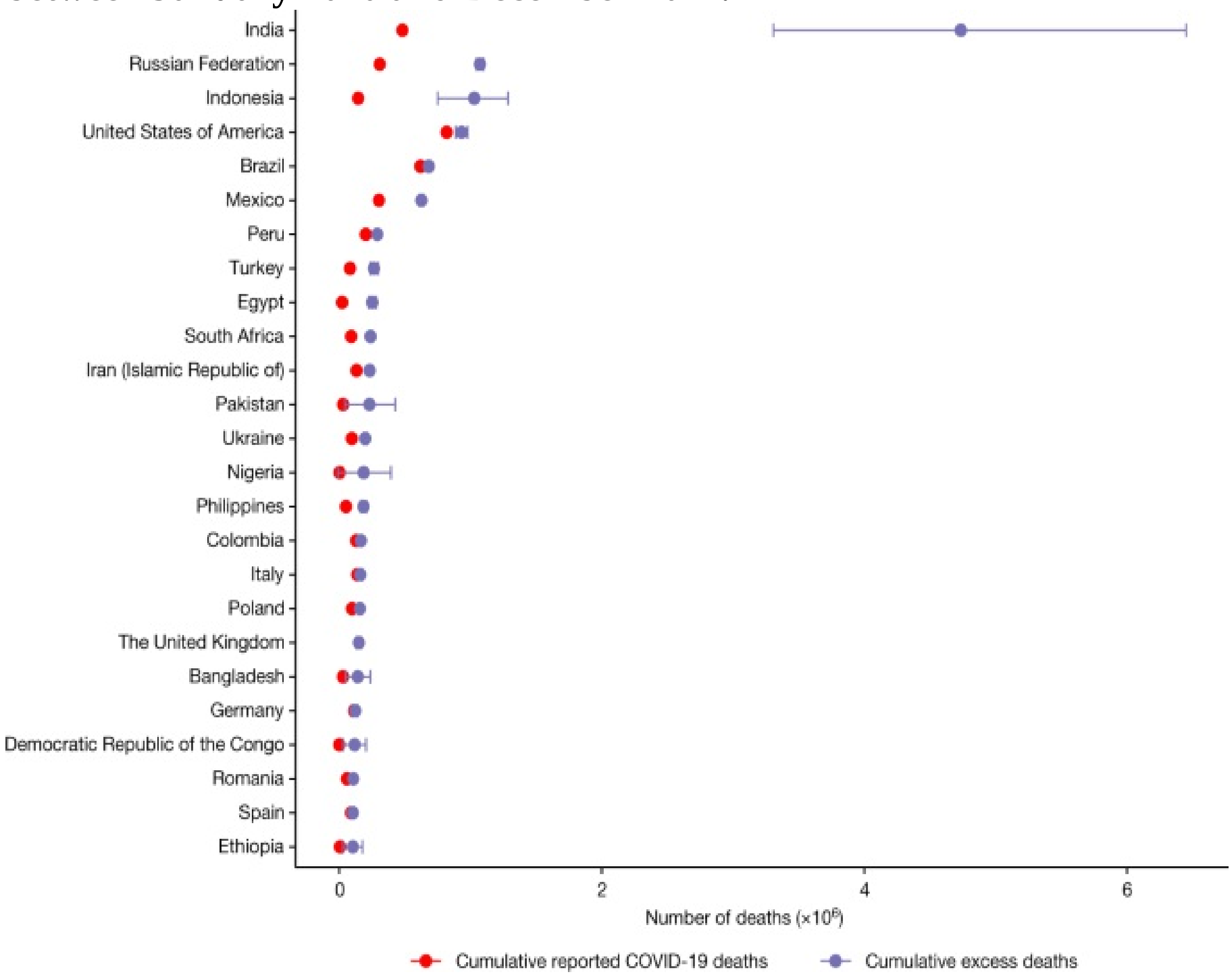
Source data

Selected country summary

Figure 3 displays the excess deaths estimates and reported COVID-19 death counts for the 25 countries with the highest numbers of estimated excess deaths, along with error bars for the uncertainty interval. The 20 countries with the highest excess estimates represent approximately half (48.9%) the

global population and account for over 80% of the estimated global excess deaths for the January 2020 to December 2021 period. These countries are (in alphabetical order) Bangladesh, Brazil, Colombia, Egypt, India, Indonesia, Iran, Italy, Mexico, Nigeria, Pakistan, Peru, the Philippines, Poland, the Russian Federation, South Africa, the United Kingdom, Turkey, Ukraine and the United States of America.

Fig. 3: The 25 countries with the highest total estimated excess deaths between January 2020 and December 2021.



The red dots show the total reported COVID-19 death numbers. The purple dots show the mean total estimated excess death numbers with the width of the bars showing the 95% uncertainty intervals.

Source data

There are an estimated 4.74 million (UI 3.31 to 6.45 million) excess deaths for India alone in the period January 2020 to December 2021, followed by 1.07 million (UI 1.05 to 1.10 million) excess deaths in the Russian Federation, 1.03 million (UI 0.75 to 1.29 million) excess deaths in Indonesia and 932K (UI 887K to 978K) excess deaths in the United States of America. Figure 3 contains countries of all possible data types in our model:

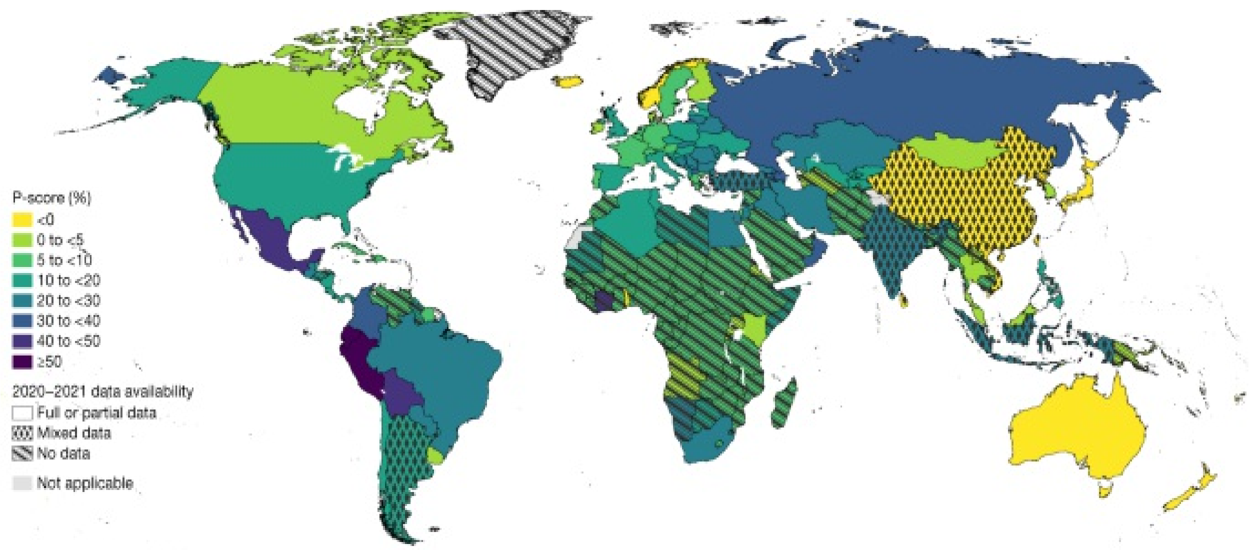
- Countries with full ACM data for the entire analysed period (Russian Federation, United States of America, Brazil, Egypt, Spain and so on) for which the uncertainty in the estimates comes from the expected (counterfactual) number of deaths and is therefore relatively narrow.

- Countries with mixed ACM data (India, Indonesia, Turkey), that is, subnational data projected to the national level. This produces an additional layer of uncertainty from the projection of the information to the national level.

- Countries for which ACM data were unavailable (Pakistan, Nigeria, Ethiopia and so on), where the estimates are derived from the covariate prediction model. In these countries the uncertainty is the highest and results are to be interpreted with the greatest caution. However, the estimated undercount in these countries is well within the plausible range of undercounts as estimated for countries with full ACM data.

The total estimated excess deaths are heavily determined by the population size of each country. Figure 4 shows a map of the estimated P-scores for 2020–2021, and Fig. 5 displays the 25 countries with the highest estimated P-scores. In both figures the P-scores are calculated using expected numbers over 2020–2021.

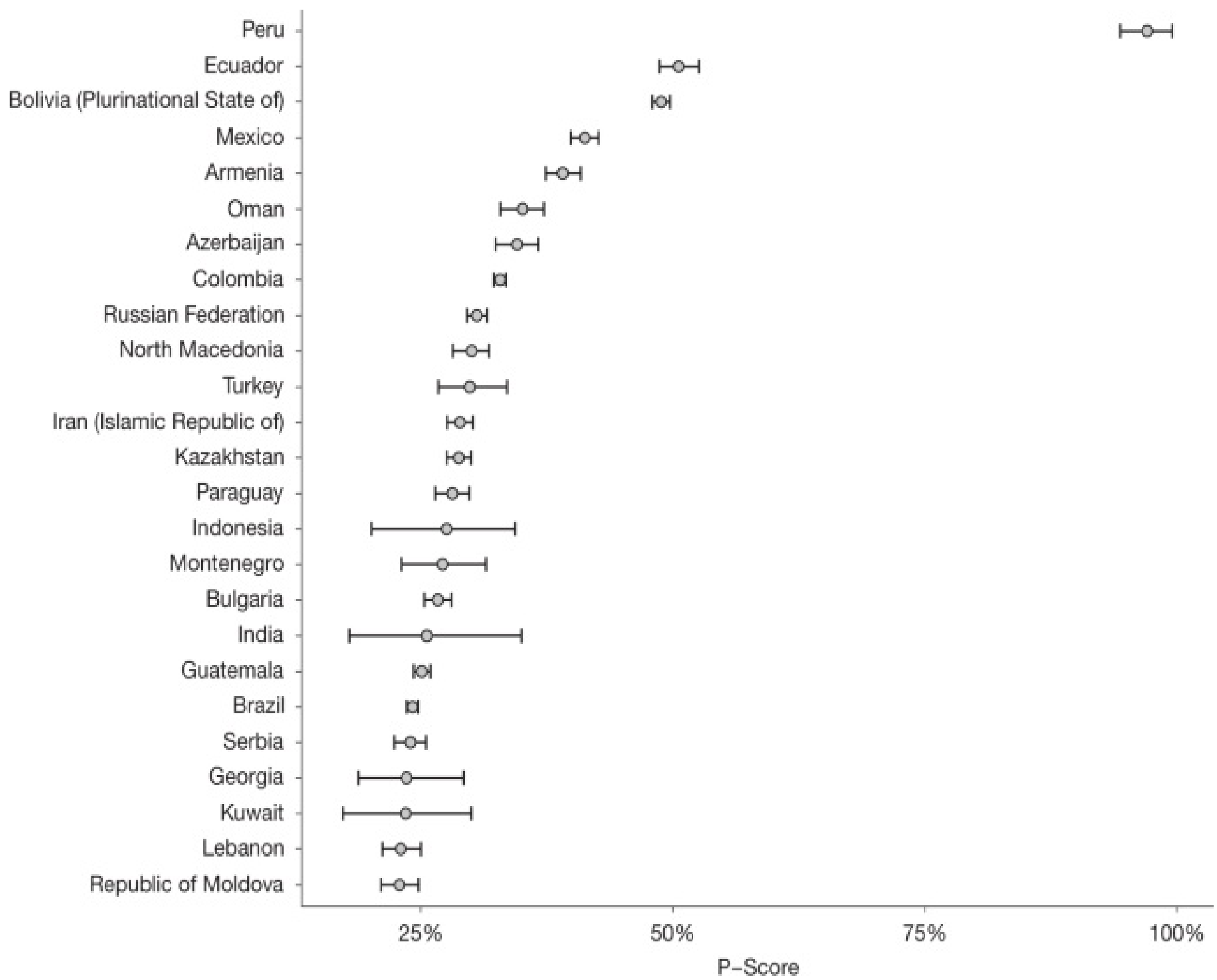
Fig. 4: Mapping estimated P-scores (excess deaths relative to expected deaths).



The map shows the geographic distribution of the mean P-scores for years 2020 and 2021 across all 194 WHO member states. The darker the colour the higher the estimated mean P-score. The patterns indicate the quality of the all-cause mortality data that were available for each respective country with the solid pattern showing full or partial data, dots for mixed data and diagonal lines for no data.

Source data

Fig. 5: The 25 countries with the highest mean P-scores (excess deaths relative to expected deaths).



The plot shows the 25 countries with the highest mean P-scores for years 2020 and 2021 after ranking all WHO member states with populations greater than 200,000 by mean P-score from highest to lowest. The mean for each country is shown using the central grey dots and the widths of the bars show the 95% uncertainty intervals.

Source data

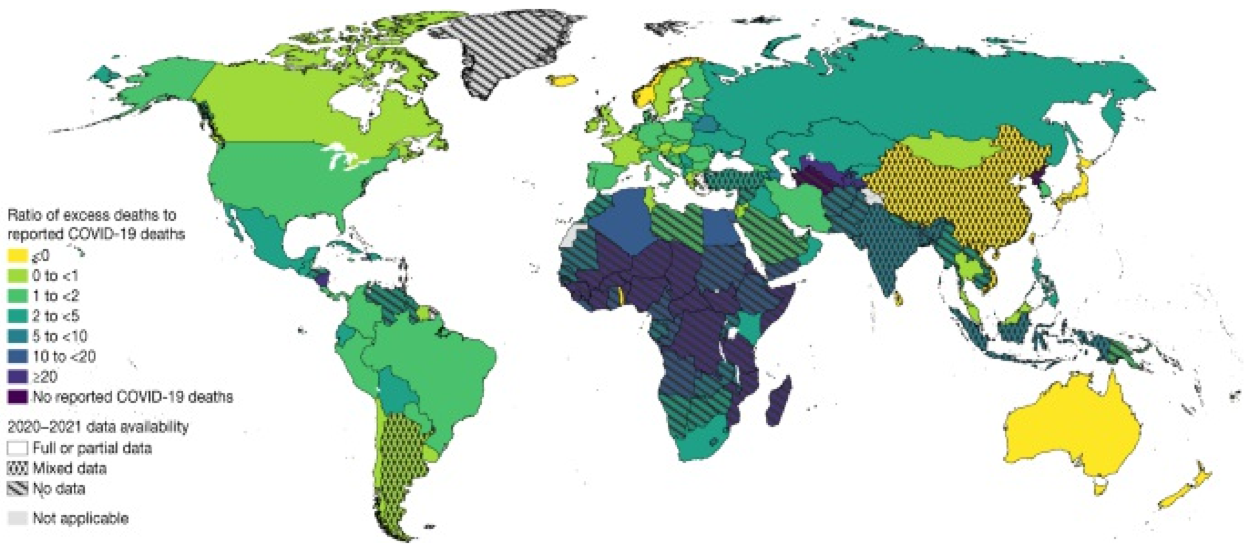
From the P-score point of view, the worst impacted countries generally have smaller populations than those from Fig. 3. For example, India and the United States of America have the highest and fourth highest estimated excess deaths in the world in absolute terms but the United States of America is absent from the top 25 and India is 21st in the list. On the other hand, smaller countries such as Peru, Bulgaria and Bolivia have been impacted more heavily, relative to population size, when examining excess mortality in

absolute terms.

So far as individual countries are concerned, Peru has a devastating P-score estimate of 97%, a doubling of deaths over the pandemic relative to what was expected. Other countries with a large increase include Ecuador with a 51% increase in deaths and Bolivia with a 49% increase.

Figure 6 maps the ratio of excess deaths to reported COVID-19 deaths. There is a wide range for this excess measure, with many countries in the AFRO region having high ratios, and countries in Western Europe having ratios closer to 1 (with some, such as France, having values below 1). Globally, over January 2020–December 2021, there were 5,420,534 reported COVID-19 deaths, and according to our estimates, the ratio of excess to reported COVID-19 deaths is 2.74 (UI 2.44 to 3.06), which is a huge discrepancy.

Fig. 6: Mapping the ratio of total excess deaths to total reported COVID-19 deaths.



The map shows the geographic distribution of the mean ratio of the total excess deaths to total reported COVID-19 deaths for years 2020 and 2021 across all 194 WHO member states. The darker the colour the higher the estimated mean ratio. The patterns indicate the quality of the all-cause mortality data that were available for each respective country with the solid pattern showing full or partial data, dots for mixed data and diagonal lines for no data.

Source data

As mentioned previously, we estimate the highest cumulative excess death numbers for India, accounting for 4.74 million deaths with a 95% credible interval of (3.31, 6.45) million. We base this estimate for India on subnational data—we have mortality data for 17 states and union territories (out of 36) over the pandemic. Using a proportionality assumption as discussed in the Methods and in Knutson et al.[15], we produce a national estimate. In the Supplementary information we summarize excess estimates from other studies that use different data sources and illustrate that our estimates are consistent with these estimates.

A crucial component of the excess calculation is the estimation of the expected number of deaths. There are two elements to the calculation, the mortality data upon which it is based and the model that is adopted. First, with respect to the data, the WHO adjust the raw mortality counts, if there is perceived to be any incompleteness in reporting (and the scaling value may be carried forward to the pandemic period). We note that as part of the process to produce excess estimates, country consultation is carried out, in which the adjusted country numbers are shared with nationally nominated focal points who are tasked with reviewing the adjusted counts. Second, for the expected counts modelling, we used splines both for the annual trend and for the within-year seasonal variation. A country for which the completeness adjustment and spline modelling provided a less than satisfactory excess estimate was Germany. Under the default data process/spline modelling the excess estimate was 195,000 with 95% credible interval (161,000, 229,000). However, on closer examination this excess estimate was too high because of a combination of data/model issues. For Germany, ACM in 2016–2018 were scaled up owing to the completeness assessment, which lead to a dip in the ACM sequence in 2019. The annual spline fit to these adjusted data produced expected numbers that were too low (and therefore an excess that was too high). Hence, we reanalysed the Germany data with unadjusted data and a linear term, rather than a spline. This produced a more realistic excess estimate of 122,000 with a 95% credible interval of (101,000, 143,000).

For Sweden, we were concerned there were similar issues because of an unnecessary completeness adjustment of the raw mortality figure reported to

the WHO in 2019 (the mortality count was lower than recent counts). The original excess estimate was 11,300 (9,900, 12,700). On closer scrutiny, we decided that this adjustment was not necessary and we redid the analysis for Sweden, again including a linear term for the annual trend instead of a spline (for the same reasons as described for Germany). This operation resulted in an estimate for Sweden that was higher, specifically 13,400 (11,700, 15,200). The changes in the excess estimates for Germany and Sweden do not change the global or EURO figures substantively (and the figures quoted in this paper are based on the revised estimates). As a side note, for both these countries when using the unadjusted data, both the linear and spline annual trend models produced similar excess estimates. However, using a spline for the annual trend can lead to sensitivity to the last year of pre-pandemic data, and a priority going forward is to systematically compare and evaluate different models for producing the expected numbers, building on recent work[16]. For the next round of estimates we will also revisit the under-reporting adjustment procedure. More details on revisiting the Germany and Sweden estimates are in the Supplementary information.

A natural inclination is to rank countries in terms of one of the metrics we have discussed. Basing rankings on point estimates will often be misleading because, particularly in the context of estimating excess mortality, there will often be considerable uncertainty in the chosen metric for a given country.

In the Methods, we include an illustrative example of a rankings analysis, using the excess rate. We show how one can evaluate the (posterior) probability that the excess is greater in one country than in another and provide graphical summaries to aid in interpretation. We also discuss the temporal comparison of rankings. In the Supplementary information, we describe a more substantive analysis that addresses how the United Kingdom compared to countries of the European Union, in terms of the excess rate and the P-score, over the pandemic.

In general, one can rarely simply look at the excess rate and associated rankings and make statements concerning the manner in which a country dealt with the pandemic, as there are many factors at play. These include: the age structure of the population, the population density and cultural practices, the government responses during different periods of the epidemic, how the

population responded to government actions, and the infectiousness and fatality rates of the various variants that were present at different times.

Discussion

These estimates call our attention to four important points. First and foremost, COVID-19 has resulted in marked global excess mortality: 14.83 million deaths (13.23, 16.58) over 2020–2021. In 2020, the excess was 4.47 (3.91, 5.07) million and in 2021, the excess was 10.36 (9.06, 11.97) million. The majority of the countries in the world have seen substantive increases in mortality.

Second, both the reported and excess estimated toll of COVID-19 were heavier in 2021 than in 2020. Third, excess mortality is much higher than reported COVID-19 mortality globally. In many countries COVID-19 deaths have been reported accurately, yet in others, the estimated excess mortality is much higher than reported COVID-19 mortality, occasionally by several orders of magnitude. In total, the estimates show that global excess mortality is 2.74 times higher than reported COVID-19 deaths. Finally, for almost half the countries of the world, tracking excess mortality is not possible using the data that are available and for these we must rely on statistical models. Critically, the missing countries are not randomly spread across the globe and so the countries with missing data may be systematically different from the sample of countries for which we do have data when we account for the age distributions of the populations, the underlying disease burdens of the countries, what is known about the strengths of their health systems and potentially most importantly, when, and how the pandemic evolved within their borders. These differences considered, some of the estimates and uncertainty intervals for locations for which we know very little must be interpreted with caution.

Despite its strengths, this study has important limitations extending beyond the data paucity mentioned above. The non-COVID counterfactual trend is derived using historical data and is sensitive to the assumptions made in the forecast. The weight given to recent data years relative to those further back in time, and how smoothly changes over time are projected to persist into the pandemic period, can influence the expected levels. We used spline models

as the basis for the modelling of the expected numbers, but as we have mentioned will revisit this choice for the next round of estimates, as such models can produce inappropriate extrapolations. Another limitation relates to the quality of the input data used. The completeness of observed deaths and the consistency in quality as well as construct validity for the various covariates used in the regression model, impact the accuracy of the empirical estimates of excess and the robustness of the predictive model for deriving excess for countries in which deaths have not been observed[16]. Currently, completeness of reporting during the pandemic is based on historic completeness of registered deaths relative to the WHO Global Health Estimates (GHE)[17]. The GHE uses multiple data sources to generate country-, age-, sex-, year- and cause-specific mortality estimates, which in turn undergo extensive country consultation. However, comparing 2019 estimates of the GHE to registered deaths from even high-quality registration systems, some differences are observed. Subsequent iterations of the excess mortality work will include updates to how the expected deaths are derived and how completeness of reporting is calculated and extrapolated to the pandemic period.

The greater proportion of global excess is derived using observed mortality data and the estimated toll is staggeringly high. To place these estimates in context, the leading cause of death in 2019 was ischaemic heart disease, with 8.9 million deaths (https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death). Information on the leading causes of death is not currently available for the pandemic years, but we would expect COVID-19 to be among the leading causes of death in 2020 and the leading cause of death in 2021. We estimate that the mean global per capita excess mortality rate was 0.06% in 2020, more than doubling to 0.13% in 2021. This surpasses the influenza pandemics of 1957, 1968 and 2009 (estimated at 0.04%, 0.03% and 0.005%, respectively)[18]. However, the 1918 influenza pandemic was magnitudes higher, with an estimated 1.0% per capita excess mortality rate, or 75 million global excess deaths when adjusted to the 2020 population[18].

Excess mortality quantifies the increase in mortality from all causes, including direct COVID-19 deaths, indirect COVID-19 deaths (for example, health-system overload) and strictly non-COVID-19 deaths (for example, those resulting from other health shocks such as violent conflict or

disasters). These estimates from the WHO cannot quantify the relative importance of each of these factors. However, considering the low levels of excess mortality in countries in which COVID-19 transmission, infection and mortality rates were low during some of the analysed period (for example, Malaysia, Mongolia, Uruguay in 2020) or its entirety (for example, Australia, Japan, New Zealand), suggests that in many countries the greater proportion of excess deaths can be attributed to COVID-19 directly. In fact, where accurately quantified, excess mortality may provide a reliable lower-bound on COVID-19 deaths considering that for several countries, we have mortality deficits or negative estimates for certain months. The greater number of these countries have high-quality reporting systems and this deficit is due to deaths from non-natural and natural causes decreasing during the analysed period[13] and there having been less severe influenza seasons in 2020 and 2021 relative to previous years[19]. In many such countries, mortality improvements may be attributable to health systems being especially geared up to respond, the populations seeking early attention because of heightened sensitivity to health issues, compliance with public health and social measures that could reduce transmission of other infections and containment measures such as lockdowns. However, we are also aware that there have been non-COVID-19-related crises that have been experienced by some countries, for example, the conflict in Armenia, for which only a proportion of the excess mortality in 2020 would rightly be associated with COVID-19.

The alternative to excess mortality estimates—that is, relying on reported COVID-19 deaths—represents a severe undercount of the toll. Indeed, even in countries with ACM data for which the estimates are much more certain, mortality has risen substantially such that excess mortality is much higher than reported COVID-19 deaths, whether it is by 50% or by several hundred per cent. There is very little chance that the countries for which ACM data are not available have been able to report COVID-19 deaths accurately. There is ample, albeit preliminary evidence from many countries for which ACM data were not obtained (such as Pakistan and Haiti), that mortality has increased, and excess mortality is much higher than reported COVID-19 deaths. As another example, in the Africa region for which ACM data are most lacking, recent studies have highlighted the underestimation in the reported statistics[20]. In this region, the direct estimates of the ratio between excess mortality and reported COVID-19 deaths spans the gamut from about

1 in Tunisia to 2.62 in South Africa, 12 in Egypt and over 20 in Algeria. Thus, although a total regional estimate of a ratio of 8.03 might seem high at first glance, it is well within bounds of the more certain ratios and the true, yet unknown ratio, may very well be even higher both regionally and globally. As already noted, the excess estimates for sub-Saharan Africa are the least robust because of a paucity of data.

In the two years within which the COVID-19 pandemic has severely impacted humanity, important lessons remain to be fully documented and harnessed as part of the global public health surveillance capacity. First, the urgent need to improve data and health information systems and the way data are collected, analysed, shared and reported. Second, the required alignments of communicable disease surveillance with the continuous strengthening of health information systems and their integration with other existing routine surveillance systems, and with demographic and geographic monitoring systems to facilitate timely and targeted interventions. COVID-19 surveillance must also be combined with Universal Health Coverage and the International Health Regulations monitoring and related indicators for health-system preparedness, including vaccine coverage and water, sanitation and hygiene services.

As shown in the Supplementary information, there is a more than doubling of excess deaths when comparing 2021 to 2020. Despite the advances in diagnostics and therapeutics in 2020 and the rapid development of vaccines throughout the year, the end of 2020 saw the permeation of the virus into highly populous societies that had previously suffered limited exposure. In 2021, the rise in infections outpaced the roll-out of vaccines in many such locations and this either led to or was worsened by the emergence of more infectious, higher fatality, SARS-CoV-2 strains such as the Delta variant. We can speculate about how vaccine hesitancy, premature relaxation of containment measures and a global COVID-19 ‘fatigue’ contributed to how the pandemic developed but this is an important area for further study. And although the variants of SARS-CoV-2 continue to emerge (https://www.who.int/activities/tracking-SARS-CoV-2-variants) and are sequenced to identify variants that are likely to produce more serious illness, there remains limited information in real time to track them to create early warning systems of global import. This is critical as even the impact of less

severe variants such as Omicron cannot be discounted among unvaccinated and older adult populations[21].

To emerge from this crisis, the world needs to be able to monitor mortality and morbidity with real-time, reliable and actionable data. Strengthened country capacity for data and information requires collaboration across governmental and non-governmental institutions, including ministries of health and finance, hospitals, insurance companies, charities, national statistics institutions, offices of the registrar general, local and regional government, think tanks, academia and more. Monitoring systems of specific causes and ACM at the national, regional and global levels may serve as an early warning system for future health emergencies that will allow more timely responses to prevent local outbreaks from inflicting harms on lives and livelihoods in their immediate surroundings and across the world. Gaps in knowledge and data lead to gaps in response. It is thus vital for future responses that countries have well-functioning CRVS systems, which are the foundation upon which monitoring, prevention and future advancements on health rest.

WHO has made all of the results of the excess mortality estimates publicly available in an interactive web application at https://worldhealthorg.shinyapps.io/covid19excess/. This tool allows transparent exploration of estimates from the country level up to the regional and global levels. All data and code are available at https://github.com/WHOexcessc19/Codebase, so that our analyses and the results are completely reproducible.

The WHO excess mortality model is a live model that will be periodically updated given additional mortality data as well as data on covariates of relevance. We will also continue to improve the statistical framework and model. These estimates also serve as inputs to other important projects such as WHO's GHE[17] and the UN's World Population Prospects[14]. Although WHO has made preliminary results available disaggregated by broad age groups and by sex for every country, region and the world and documented the method to do this[22], this is work in progress and hence will be reported in a future paper.

Methods

Process

The process for producing the estimates of excess mortality consisted of three main steps. First, a Technical Advisory Group (TAG) was established to develop a set of methods that were used to produce estimates of excess deaths associated with the COVID-19 pandemic in countries. Second, WHO member states were consulted on the estimates, input data sources and methods. Finally, feedback from the countries was then incorporated into the modelling to update the estimates. The details of each step are described below.

In February 2021, the WHO, in collaboration with the United Nations Department of Economic and Social Affairs, formed the TAG on COVID-19 Mortality Assessment to advise on the development of analytical methods for estimating excess mortality in all countries. The TAG is composed of leading demographers, epidemiologists, economists, data and social scientists and statisticians from a range of backgrounds and geographies. A complete list of the TAG members is provided at the end of the paper. In addition to determining the levels and the age and sex distributions of the excess deaths associated with the COVID-19 pandemic, the expertise of the TAG has been leveraged to study the impact of the pandemic on broader areas such as inequality in COVID-19 mortality between and within countries, death registration and reporting systems, and how existing surveys and censuses can be used to fill in data gaps to quantify the impact of the pandemic. At the time of writing, this work is still ongoing.

In August 2021, a circular letter was sent to all WHO member states to nominate focal points to take part in country consultation. Member states were requested to review and provide feedback on the preliminary estimates of COVID-19 excess mortality and submit additional data that may not have been previously available to WHO. The first round of the country consultation was conducted between October and November 2021 through WHO’s Country Portal, an online platform to facilitate data exchange between member states and WHO, for which the draft estimates and

methodology for each country were made available to the designated national focal points. Countries that had not nominated a focal point were approached through their respective WHO country office or permanent mission in Geneva, Switzerland.

Between October 2021 and February 2022, a global technical consultation and two information sessions with member states were held to brief them on the progress and exchange views on the methodology. A series of regional webinars and technical consultations with individual countries were also organized for further discussion on input data, methods and estimates. By the end of March 2022, 140 countries (or 72% of the 194 member states) had participated in the country consultation, 65 had provided some data and 76 had provided feedback, which was then used to generate updated estimates. The revised estimates for a 24-month period from January 2020 to December 2021 were shared with the national focal points in March 2022.

The process of generating the estimates of excess mortality associated with the COVID-19 pandemic has followed the Guidelines for Accurate and Transparent Health Estimates Reporting[23]. In view of the fast-changing situation surrounding the pandemic, the excess mortality estimates will continue to be refined and revised as more data are identified and the methodology evolves over time.

Data

The specific countries from each region for which data have been gathered are listed in the Supplementary information and are shown in Extended Data Fig. 1. Estimates of the excess mortality associated with the COVID-19 pandemic require historical ACM data that can be used to generate the death numbers under a hypothetical non-COVID-19 scenario, as well as ACM data for the target years against which the counterfactual is contrasted to calculate the excess. In the absence of nationally representative data, subnational data can be used to estimate national totals.

Reported ACM data at the national level on a weekly or monthly basis are available for only a subset of countries. The data used in this study span multiple sources:

- Data routinely shared with WHO as part of its standing agreement with member states as well as specifically provided to WHO in response to a data call for this project.
- Data that have been reported by European countries to Eurostat according to the European Statistical System[24].
- Data that have been compiled for the Human Mortality Database as part of the Short-term Mortality Fluctuations project[25,26].
- Data that have been compiled in the World Mortality dataset[13].

Additionally, annual level data for 2020 and/or 2021 were obtained from the national statistics offices of China[27,28], Grenada[29], Saint Kitts and Nevis[30], Saint Vincent and the Grenadines[31], Sri Lanka[32] and Vietnam[33].

The countries with current reported ACM generally have ACM data for the pre-pandemic period as well. For those without such historical data, the WHO GHE[34] database was used. Using the annual historic mortality we forecasted expected ACM to 2020 and 2021, to provide the expected mortality in these locations. The method for this forecast will be described shortly.

In addition to the data on reported ACM and the estimates from the GHE, the final dimension to the input data are variables that can potentially be used as predictors for excess mortality in those countries/time periods without ACM data. The strategy applied to create a covariate list was pragmatic and focused on identifying those variables that have been found to be contextually important and that have been measured/estimated in the majority of countries. The predictor variables are composed of both time-varying and time-invariant variables. Time-varying variables were the test positivity rate, temperature, confirmed COVID-19 death rate per 100,000 population (which is reported to the WHO), COVID-19 positive test rate per 100,000 population (from Our World in Data https://ourworldindata.org) and a variable constructed from a number of containment measures[35]. The COVID-19 death rate and the positive test rate are available for all member states for the entire period (https://covid19.who.int/). The time-invariant variables were a binary

measure of the income level (low/middle versus high) and the historic diabetes prevalence and cardiovascular mortality rates as estimated by the Global Burden of Disease project[9].

Subnational-level (states, provinces, cities, collections thereof and so on) data were obtained from various sources for Argentina[36], India[37,38,39,40,41,42], Indonesia[43] and Turkey[44].

Statistical models

We write the excess in country c at time t as

$${\delta }_{c,t}={Y}_{c,t}-{E}_{c,t},$$
(1)

where ${Y}_{c,t}$ is the realized ACM and ${E}_{c,t}$ is the ACM that would be expected in the absence of the pandemic. Even for countries with fully observed ACM during the pandemic the excess is a random quantity, because we do not know the counts ${E}_{c,t}$ that would have occurred in the absence of the pandemic—the latter is the result of a modelling exercise, which produces forecasted ACM, with associated uncertainty.

The major challenges for modelling are to form a coherent approach in the face of disparate data sources of varying degrees of quality and in different spatially and temporally aggregated forms. We constructed a model from first principles within a Bayesian inferential framework, and as a first step developed a framework in which we directly model the raw death counts (as opposed to derived quantities such as rates). Death is binary, and so must follow a Bernoulli distribution, and it is also statistically rare, and so the Bernoulli can be accurately approximated by a Poisson distribution. The advantage of the latter is that it is amenable to manipulation when one considers subsets of availability such as over space (when subnational data only are available) or over time (when annual counts only are available). In the Poisson model the variance equals the mean, which is restrictive as mortality data typically exhibit greater variability than the nominal variance. Hence, we use models that allow for such excess-Poisson variation.

For modelling all countries of the world we need to consider various data situations. Although some countries have full data, others have annual or subnational data only, and for countries with no data we need to build a predictive model based on country-specific variables. Extended Data Figure 2 shows the relationship between the different models, and how they feed into the excess calculation.

Model for expected numbers

For all countries and time points we model the expected numbers on the basis of historic data (for most countries, the period 2015–2019 was used for this modelling). We use a negative binomial model that allows for excess-Poisson variation. The annual historic yearly trend in ACM is modelled using a spline model, and within-year variation using a seasonal spline model. A spline is a flexible approach to modelling that allows departures from a linear association[45]. A negative binomial model has two parameters, a mean (which is obtained from spline components), and a scale parameter that accounts for excess-Poisson variation. The mean count for country c and in month t is modelled as:

$${{\rm{M}}{\rm{e}}{\rm{a}}{\rm{n}}{\rm{c}}{\rm{o}}{\rm{u}}{\rm{n}}{\rm{t}}}_{c,t}=\exp ({{\rm{a}}{\rm{n}}{\rm{n}}{\rm{u}}{\rm{a}}{\rm{l}}{\rm{t}}{\rm{r}}{\rm{e}}{\rm{n}}{\rm{d}}}_{c}+{{\rm{s}}{\rm{e}}{\rm{a}}{\rm{s}}{\rm{o}}{\rm{n}}{\rm{a}}{\rm{l}}{\rm{c}}{\rm{o}}{\rm{m}}{\rm{p}}{\rm{o}}{\rm{n}}{\rm{e}}{\rm{n}}{\rm{t}}}_{c,t})$$

where the annual trend uses a thin-plate spline and the seasonal component uses a cyclic cubic spline. After fitting to pre-pandemic data, we project the modelled trend forward to predict expected counts, by month, for 2020 and 2021. There is uncertainty in these predictions, which we incorporate into the excess mortality uncertainty intervals we produce.

For some countries, we only have national historic ACM data. For such countries we model within-year variation using temperature as a surrogate for seasonality. Full details of all modelling steps are given in Knutson et al.[15].

Model for countries without full pandemic data

For almost half of the 194 WHO member states we do not have the ACM counts over the pandemic, and so must predict them using country characteristics. We choose a simple form for this prediction model, with mean

$$\text{E}\left[{Y}_{{ct}}{\rm{|}}{E}_{c,t}\right]={E}_{c,t}{\theta }_{c,t},$$
(2)

where $({\theta }_{c,t} > 0)$ is a relative rate parameter. If $({\theta }_{c,t} > 1)$, then for country c and at month t the mortality is greater than expected, whereas if $({\theta }_{c,t} < 1)$, then for country c and at month t the mortality is less than expected. We used G time-invariant variables, $({Z}_{{gc}})$ (these are annual values from 2019). These were an indicator of high income and/or low or middle income, the cardiovascular mortality rate in 2019 and the diabetes prevalence rate in 2019. In addition, we used B time-varying variables: a containment variable (it is calculated using all ordinal containment and closure policy indicators and health-system policy indicators, for further details see Hale et al.[14]), the square root of the reported COVID-19 death rate, temperature and the COVID-19 test positivity rate. We then build a log-linear model for the rate parameter:

$$\log {\theta }_{c,t}=\mathop{\underbrace{\alpha }}\limits_{{\rm{I}}{\rm{n}}{\rm{t}}{\rm{e}}{\rm{r}}{\rm{c}}{\rm{e}}{\rm{p}}{\rm{t}}}+\mathop{\underbrace{\mathop{\sum }\limits_{g=1}^{G}{\gamma }_{g}\,{Z}_{gc}}}\limits_{{\rm{T}}{\rm{i}}{\rm{m}}{\rm{e}}-{\rm{i}}{\rm{n}}{\rm{v}}{\rm{a}}{\rm{r}}{\rm{i}}{\rm{a}}{\rm{n}}{\rm{t}}\,{\rm{c}}{\rm{o}}{\rm{n}}{\rm{t}}{\rm{r}}{\rm{i}}{\rm{b}}{\rm{u}}{\rm{t}}{\rm{i}}{\rm{o}}{\rm{n}}{\rm{s}}}+\mathop{\underbrace{\mathop{\sum }\limits_{b=1}^{B}{\beta }_{bt}{X}_{bct}}}\limits_{{\rm{T}}{\rm{i}}{\rm{m}}{\rm{e}}-{\rm{v}}{\rm{a}}{\rm{r}}{\rm{y}}{\rm{i}}{\rm{n}}{\rm{g}}\,{\rm{c}}{\rm{o}}{\rm{n}}{\rm{t}}{\rm{r}}{\rm{i}}{\rm{b}}{\rm{u}}{\rm{t}}{\rm{i}}{\rm{o}}{\rm{n}}{\rm{s}}}+\mathop{\underbrace{{{\epsilon }}_{c,t}\,}}\limits_{{\rm{E}}{\rm{x}}{\rm{c}}{\rm{e}}{\rm{s}}}$$

{\rm{s}}-{\rm{P}}{\rm{o}}{\rm{i}}{\rm{s}}{\rm{s}}{\rm{o}}{\rm{n}}\,{\rm{v}}{\rm{a}}{\rm{r}}{\rm{i}}{\rm{a}}{\rm{t}}{\rm{i}}{\rm{o}}{\rm{n}}}$$
(3)

where ${\rm{\exp }}\left({\gamma }_{g}\right)$ and ${\rm{\exp }}\left({\beta }_{{bt}}\right)$ denote relative rate parameters and ${{\epsilon }}_{c,t}\sim {\rm{N}}(0,{{\sigma }}_{{\epsilon }}^{2})$ are independent error contributions that pick up random variation unexplained by the log-linear regression function. The time-varying coefficients allow the associations to evolve during the pandemic. As we desire the evolution to be smooth in time, for these time-varying coefficients ${\beta }_{c,t}$ we use a random walk of order 2 (RW2) prior that encourages smooth estimates[46]. In equation (2) above, we have conditioned on known expected numbers. In reality, and as just described, these are modelled to give a distribution over plausible values. The uncertainty in the expected predictions ${E}_{c,t}$ is well modelled by a gamma distribution, and the advantage of this choice is that it can be conveniently combined with a Poisson model to produce a negative binomial model with the log-linear mean given by equation (3). Full details (including evidence of the accuracy of the gamma model) can be found in Knutson et al.[15].

This model was fitted to all countries with observed monthly ACM data over some portion of 2020–2021, using the integrated nested Laplace approximation method[47], to obtain posterior distributions over the unknown parameters. The resultant posterior distribution reflects the uncertainty in the parameters (both in the expected numbers and the log-linear covariate model), and can be used to construct a predictive distribution for the ACM in countries with no data or partial data.

For some countries, only subnational data were available, and so we construct a model for the national ACM data using a proportionality assumption, expanding on previous work[48]. We describe the model in the context of India. We use ACM data from 17 states and union territories out of 36 (data from different numbers of states are available in different pandemic months) to infer the national total, under the assumption that the proportion of deaths in the states with available data remains approximately constant over time. For

example, if a state historically accounts for 10% of deaths in India, one would predict a national death total of 10$\times$ the observed number of deaths in that state only. Under the Poisson framework, this proportionality assumption yields a multinomial distribution for the fractions of deaths and we can predict the unknown national totals over the course of the pandemic after fitting the multinomial model.

Extensive model validation was carried out for both the countries with no data, and those with subnational data only. This included exercises in which we systematically removed all data for each country in turn, or we removed data for all countries for single months. We then predicted these removed data using the retained data and evaluated model performance using metrics such as bias and the coverage of prediction intervals. Results for these exercises can be found in the supplementary materials of Knutson et al.[15]. We emphasize that the model (3) is not used for countries with subnational data.

For other countries that have annual (but not monthly) national data during the pandemic, we lean on the fact that the distribution of Poisson monthly counts, given the annual count, is multinomial with probabilities that are the normalized rate parameters, that is,

$${p}_{c,t}=\frac{{E}_{c,t}{\theta }_{c,t}}{\mathop{\sum }\limits_{{t}^{{\prime} }=1}^{12}{E}_{c,{t}^{{\prime} }}{\theta }_{c,{t}^{{\prime} }}}.$$

where the rates ${\theta }_{c,t}$ are defined via the log-linear covariate model (3). This gives us a way to apportion the annual counts to the constituent months.

Our approach, differs from those of the other two global endeavours of the Institute for Health Metrics and Evaluation (IHME)[49] and *The Economist*[50]. We have used a very conventional statistical modelling approach in which a parametric model is fitted using Bayesian inferential machinery, and with the models for different data types being consistent with each other to make the country by country results directly comparable to each other. As an example, if the mortality in subnational regions are Poisson random variables, then the sum (the mortality in the country) is also Poisson. Further, given the total mortality in a country the subnational counts follow a multinomial

distribution. Our framework exploits these relationships when we formulate models for the situation in which we have subnational data only. Similarly, our annual model (for countries with such data only) is consistent with the monthly models we use for the majority of the countries. The IHME approach is unprincipled and not transparent and corresponds to a number of steps being bolted together, without a coherent model tying them together. Rather than using a direct count model based on a Poisson framework, the IHME approach models the log of the excess rate as a function of covariates, without any weighting, so that the population sizes of the different countries do not feed into the uncertainty calculation. A fundamental problem with the overall approach is that the uncertainty intervals are constructed in a non-standard and ad hoc way, so that the confidence intervals, in particular, will not be accurate representations of the true uncertainty. *The Economist* approach models the excess rate with a flexible tree-based machine learning technique, gradient boosting. The approach is clearly described and uses a resampling technique, the bootstrap, to form interval estimates, but there is no theory to support the use of the bootstrap with boosting, and so again, the uncertainty intervals should be viewed sceptically. A full description and critique of the alternative methods are available in Knutson et al.[15]. In the Supplementary information, we provide a comparison between point and interval country estimates obtained by the methods of the WHO, IHME and *The Economist*.

P-scores

Recall that the P-score is defined as the ratio of the excess to the expected, expressed as a percentage. Mathematically, this corresponds to,

$${\text{PS}}_{c,t}=100\times \frac{{Y}_{c,t}-{E}_{c,t}}{{E}_{c,t}},$$

and ${\text{PS}}_{c,t}\ge -100,$ with zero deaths corresponding to –100, negative values corresponding to fewer deaths than expected and larger positive values corresponding to increasing levels of relative excess mortality. Under the model (2), we have $\text{E}\left[{Y}_{c,t}\right]={E}_{c,t}{\theta }_{c,t}$ so that

$$\text{E}\left[{\text{PS}}_{c,t}\right]=100\times \left({\theta

}_{c,t}-1\right).$$

For countries whose ACM is unobserved, the rate is modelled via the log-linear form (equation (3)) which gives a specific form to the manner in which we assume the P-score changes as a function of country-specific covariates.

Rankings

A natural, if sometimes unfortunate, inclination is to attempt to rank regions or countries in terms of the various metrics. Statistically, this is fraught with difficulties. The easiest approach, which is often followed in the media, is to simply rank on the basis of a point estimate of the metric, such as the mean or the median. The obvious problem with this approach is that the uncertainty in estimation is not accounted for. Using the Bayesian machinery that we use for inference we can account for the uncertainty probabilistically. In the simplest case of two countries, let ${X}_{1}$ and ${X}_{2}$ represent the excess rates in countries 1 and 2, respectively. We can then evaluate the (posterior) probability that ${X}_{1} > {X}_{2}$, and report this, rather than a binary statement that country 1 has a higher rate than country 2. The extension to multiple countries is immediate, as is the ability to calculate the probability of a higher rate in one country as compared to any collection of other countries.

For illustration of the issues of assessing rankings, we select six European countries that have overlap in their excess rate uncertainty (that is, posterior) distributions. In the left panel of Extended Data Fig. 3 we display posterior distributions for the excess rates of the six countries, ordered from top to bottom by highest to lowest median excess rate. There is clearly overlap in many of the distributions, but quantitative statements on the rankings require more than these plots. In the right panel of Extended Data Fig. 3 we present scatterplot representations of the bivariate probability distributions describing the relationships between pairs of countries. The red lines offer a reference by which we can evaluate the ranking probabilities (by calculating the fractions of points that are either side of the line). For example, the probabilities that the rate for Slovenia is greater than that of each of Italy, Estonia, Spain, the United Kingdom and Portugal are 0.546, 0.749, 0.988, 0.992 and 0.999, respectively. Even these plots do not give the complete picture as they are

two-dimensional summaries of a six-dimensional object (the probability distribution over the six rates). We can provide other summaries, for example, the probability that the rate in Slovenia is greater than the rates in all of the other five countries is 0.479.

The rankings just discussed are based on the cumulative excess rate over January 2020–December 2021. Another potentially interesting summary is the relative rankings of countries' rates over time. In Extended Data Fig. 4 we plot the excess rate over time (top panel) and the ranking probabilities (bottom panels). In each month, we calculate the probabilities that the rate of each country is highest, second highest and so on. In 2020, we see that among the six countries considered, Spain, the United Kingdom and to a lesser extent Italy, have high rates, whereas in 2021, Italy and Slovenia and to a lesser extent Spain have high rates. The rate in Estonia is generally low, apart from the last few months of 2021.

The Supplementary information contains a more substantive example where we consider the rankings of 27 countries of the European Union and the United Kingdom over time, in terms of both the excess rate and the P-score.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data used as inputs to our modelling analyses are contained in model input files located at https://github.com/WHOexcessc19/Codebase. The site also contains all generated quantities and output datasets. Source data are provided with this paper.

Code availability

All analyses were performed using R version 3.6.3. Scripts are available at https://github.com/WHOexcessc19/Codebase, which allows the estimates to

be reproduced.

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020. *World Health Organization* https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (2020).

2. Riffe, T. & Acosta, E. Data Resource Profile: COVerAGE-DB: a global demographic database of COVID-19 cases and deaths. *Int. J. Epidemiol.* **50**, 390–390f (2021).

3. Checchi, F. & Roberts, L. *Interpreting and Using Mortality Data in Humanitarian Emergencies (2005)* https://odihpn.org/wp-content/uploads/2005/09/networkpaper052.pdf (2005).

4. *The True Death Toll of COVID-19: Estimating Global Excess Mortality* https://www.who.int/data/stories/the-true-death-toll-of-covid-19-estimating-global-excess-mortality (Word Health Organization, 2021).

5. Farr, W. *Vital Statistics: A Memorial Volume of Selections from the Reports and Writings of William Farr* (Offices of the Sanitary Institute, 1885).

6. Murray Christopher, J. L., Lopez Alan, D., Chin, B., Feehan, D. & Hill Kenneth, H. Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918–20 pandemic: a quantitative analysis. *Lancet.* **368**, 2211–2218 (2006).

7. Beaney, T. et al. Excess mortality: the gold standard in measuring the impact of COVID-19 worldwide? *J. Royal Soc. Med.* **113**, 329–334 (2020).

8. *SCORE for Health Data Technical Package: Global Report on Health Data Systems and Capacity, 2020* (World Health Organization, 2021).

9. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990– 2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet.* **396**, 1204–1222 (2020).

10. Karlinsky A. International completeness of death registration 2015–2019. Preprint at *medRxiv* https://doi.org/10.1101/2021.08.12.21261978 (2021).

11. *Demographic Yearbook 2021 Technical Reports* (United Nations Statistics Division, 2021).

12. Aron J. & Muellbauer J. A pandemic primer on excess mortality statistics and their comparability across countries. *Our World in Data* https://ourworldindata.org/covid-excess-mortality (2020).

13. Karlinsky, A. & Kobak, D. Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. *eLife* **10**, e69336 (2021).

14. *World Population Prospects 2019* https://population.un.org/wpp/Publications/ (United Nations, 2019).

15. Knutson V., Aleshin-Guendel S., Karlinsky A., Msemburi W. & Wakefield J. Estimating country-specific excess mortality during the COVID-19 pandemic (2023). *Ann. Appl. Stat.* (in the press).

16. Scholey J. Robustness and bias of European excess death estimates in 2020 under varying model specifications. Preprint at *medRxiv* https://doi.org/10.1101/2021.06.04.21258353 (2021).

17. *Global Health Estimates: Life Expectancy and Leading Causes of Death and Disability* (World Health Organization, 2021).

18. Simonsen, L. & Viboud, C. Mortality: a comprehensive look at the COVID-19 pandemic death toll. *eLife* **10**, e71974 (2021).

19. Kung, S., Doppen, M., Black, M., Hills, T. & Kearns, N. Reduced

mortality in New Zealand during the COVID-19 pandemic. *Lancet* **397**, 25 (2021).

20. Cabore J. W. et al. COVID-19 in the 47 countries of the WHO African region: a modelling analysis of past trends and future patterns. *Lancet Glob. Health* **10**, e1099–e1114 (2022).

21. Subissi L. et al. An early warning system for emerging SARS-CoV-2 variants *Nat. Med.* **28**, 1110–1115 (2022).

22. *Methods for Estimating the Excess Mortality Associated with the COVID-19 Pandemic* https://www.who.int/publications/m/item/methods-for-estimating-the-excess-mortality-associatedwith-the-covid-19-pandemic (World Health Organization, 2022).

23. Stevens, G. A. et al. Guidelines for accurate and transparent health estimates reporting: the GATHER statement. *PLoS Med*. **13**, e1002056 (2016).

24. Deaths by week and sex. *Eurostat* https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Weekly_death_statistics&stable2021 (2022).

25. *Human Mortality Database. Short-term Mortality Fluctuation Data Series* (Max Planck Institute for Demographic Research and Univ. California Berkeley, accessed 29 March 2022); https://www.mortality.org/Data/STMF.

26. Islam N., et al. Excess deaths associated with COVID-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries. *BMJ* **373**, n1137 (2021).

27. National Bureau of Statistics of China. *China Statistical Yearbook 2021* (China Statistics Press, 2021).

28. National Bureau of Statistics of China. *National Economy Continued to Recover with Expected Development Targets Well Achieved in 2021* (Press Release)

http://www.stats.gov.cn/english/PressRelease/202201/t20220117_18264(
(2021).

29. Live births and deaths 2000 to 2020. *Grenada Central Statistics Office* https://stats.gov.gd/subjects/population-2/live-births-and-deaths-2000-to-2020 (2021).

30. Births and deaths. *St. Kitts and Nevis Department of Statistics* http://stats.gov.vc/subjects/vital-statistics/ (2021).

31. Vital statistics. *St. Vincent & the Grenadines Ministry of Finance Economic Planning and Information Technology* http://stats.gov.vc/subjects/vital-statistics/ (2021).

32. Population - vital statistics. *Sri Lanka Department of Census and Statistics* http://www.statistics.gov.lk/Population/StaticalInformation/VitalStatistic (2022).

33. Statistical yearbook 2020. *Vietnam General Statistics Office* https://www.gso.gov.vn/en/data-and-statistics/2021/07/statistical-yearbook-of-2020/ (2021).

34. *WHO Methods and Data Sources for Life Tables 1990–2019 Technical Report* (Department of Data and Analytics, Division of Data, Analytics and Delivery for Impact, WHO, 2020).

35. Hale, T. et al. *Variation in Government Responses to COVID-19 Technical Report* (Univ. Oxford, 2020).

36. Garbovetzky, A. Impacto poblacional de la pandemia: crecieron las muertes 26% y cayeron los nacimientos 14%. *La Voz* (14 January 2022).

37. The Hindu Data Team. ‘Excess deaths’ during the pandemic in India was 5.8 times the official COVID-19 death toll. *The Hindu* (11 September 2021).

38. Mortality - total deaths by district and state, aggregated by month and

year. *Indian Journalist Consortium & Development Data Lab* https://www.devdatalab.org/covid (2021).

39. Vital statistics of India based on the Civil Registration System 2019. *Office of the Registrar General, India* https://censusindia.gov.in/nada/index.php/catalog/42541/download/4616 (2021).

40. Births and deaths. *Government of Tamil Nadu* https://www.crstn.org/birth_death_tn/MisRep.jsp (2022).

41. Civil registrations. *Government of Kerala* https://cr.lsgkerala.gov.in/Pages/Map.php (2022).

42. Birth and death registration. *Government of Karnataka* https://ejanma.karnataka.gov.in/ (2022).

43. Data on the number of population death reports by gender per urban village in DKI Jakarta province in 2021. *Jakarta Open Data* https://data.jakarta.go.id/dataset/data-penerbitan-akta-kematian-tahun-2021 (2021).

44. Excess mortality in Turkey. *gucluyaman* https://github.com/gucluyaman/Excess-mortality-in-Turkey (2021).

45. Wood, S. N. *Generalized Additive Models: An Introduction with R* 2nd edn (CRC Press, 2017).

46. Rue H. & Held L. *Gaussian Markov Random Fields: Theory and Application* (Chapman and Hall/CRC Press, 2005).

47. Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *J. Royal Stat. Soc. B.* **71**, 319–392 (2009).

48. Karlinsky, A. Estimating national excess mortality from subnational data: application to Argentina. *Rev. Panam. Salud Publica* **46**, e19 (2022).

49. COVID-19 Excess Mortality Collaborators. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. *Lancet* **399**, 1513–1536 (2022).

50. The pandemic's true death toll. *The Economist* https://www.economist.com/graphic-detail/coronavirus-excess-deaths-estimates (25 October 2022).

51. Technical Advisory Group on COVID-19 Mortality Assessment. *World Health Organization* https://www.who.int/data/technical-advisory-group/covid-19--mortality-assessment/membership (2021).

Acknowledgements

We thank S. Narita for her support in the country consultation process and contributions to the description of that process. We also thank the Technical Advisory Group on COVID-19 Mortality Assessment[51] for their contributions. Members include: T. Adair, S. Asma, A. Badr, S. Bignami, D. Bradshaw, C. Camarda Giovanni, B. Cao, C. Castillo-Salgado, S. Chatterji, H. Chen, O. Chestnuv, J. Emina, J. Friedman, J. Garcia Arias, P. Gerland, Y. Gu, S. Helleringer, A. Hosseinpoor, N. Islam, P. Jha, T. Johnson Takwa, A. Karlinsky, A. Krishnan, N. Lai, S. Lewington, X. Lin, L. Liu, D. Ma Fat, S. Mac Feely, K. McCormack, S. Mrkic, W. Msemburi, B. Mukherjee, B. Munkombwe, S. Narita, A. Noymer, F. Perucci, O. Phillips, B. Queiroz, T. Riffe, J. Sandberg, C. Santamaría Ulloa, C. Sawyer, O. Seidu, V. Shkolnikov, I. E. Udo, E. Van Kleef, J. Wakefield, H. Wang and J. Wilmoth. We are grateful for contributions from Member States, national focal points and WHO country offices as well as the support of the following individuals at WHO regional offices and headquarters: K. Adib, Z. Ali Piazza, E. A. Aly, S. Atia, R. J. Brennan, A. Cox, H. Doctor, B. Droti, M. Duan, D. Estevez Fernandez, J. Gao, S. Garcia Saiso, A. S. Gueye, M. Jhalani, P. Jorgensen, H. C. Karamagi, H. K. Kipruto, S. Mac Feely, T. Matsui, D. Novillo Ortiz, B. Olowokure, A. Palmquist, R. Rajbhandary, A. Rashidian, R. M. Rastogi, G. Rockenschaub, E. C. Salvador, R. Sampah Bissegger, P. Soliz, and C. R. Ugarte Casafranca. The work on the WHO estimates of excess mortality associated with COVID-19 pandemic was produced under the leadership of S. Asma (Assistant Director-General, Data, Analytics and Delivery for

Impact), and I. Socé Fall (Assistant Director-General, Health Emergency Response), in collaboration with the United Nations Department of Economic and Social Affairs (UN DESA) under the leadership of L. Zhenmin, United Nations Under-Secretary-General for Economic and Social Affairs and S. Schweinfest, Director of the Statistics Division of UN DESA.

Author information

Authors and Affiliations

1. World Health Organization, Geneva, Switzerland

   William Msemburi & Somnath Chatterji

2. Hebrew University, Jerusalem, Israel

   Ariel Karlinsky

3. Department of Biostatistics, University of Washington, Seattle, WA, USA

   Victoria Knutson, Serge Aleshin-Guendel & Jon Wakefield

4. Department of Statistics, University of Washington, Seattle, WA, USA

   Jon Wakefield

Contributions

S.C., W.M. and J.W. conceptualized the project. W.M. and A.K. curated the data. W.M., A.K., V.K., S.A.-G. and J.W. performed the formal analysis. J.W. was responsible for the methodology. Project administration was provided by W.M. V.K., S.A.-G. and W.M were responsible for the analytical software. W.M, V.K. and J.W. did the visualization. W.M. and A.K. wrote the original draft. W.M., A.K., S.C. and J.W. reviewed and edited the manuscript.

Corresponding author

Correspondence to William Msemburi.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Enrique Acosta, Geert Molenberghs and Oliver Watson for their contribution to the peer review of this work. Peer reviewer reports are available.

Extended data figures and tables

Extended Data Fig. 1 Mapping the availability of all-cause mortality data.

The countries in dark blue have all 24 months of data available for January 2020 to December 2021 whereas those in purple have monthly data available but for less than 24 months. For the countries in green we only have either subnational or annual data for some or all of the period and for those in yellow do not have any representative all-cause mortality data available for the pandemic period.

Extended Data Fig. 2 Overview of modelling strategy to produce excess mortality estimates for all countries.

The flowchart outlines how the excess mortality associated with the COVID-19 pandemic is estimated. By definition, this is the difference between the all-cause mortality (ACM) during the pandemic period and that which was expected had the pandemic not occurred (light-green). Starting with the historic monthly ACM data (light-blue) we apply the Negative Binomial Spline model to generate the expected deaths for each country. The data that are available inform the modelling strategy employed to calculate the pandemic period ACM (light-pink). For the places with full data, the reported ACM are taken as is. For those countries with only subnational ACM data, a Multinomial Subnational model is employed to derive national level ACM based on the historic fractions of deaths observed in subnational regions. The overdispersed Poisson Covariate model is fitted to countries with monthly pandemic data, and this model is used to estimate pandemic period ACM for countries without any reported ACM. The covariate model is also used to infer the within-year monthly pandemic ACM in countries with only annual data, via the Multinomial Covariate model.

Extended Data Fig. 3 Excess mortality densities and ranking probabilities for select countries.

Left: Ridgeplots representing the uncertainty in the cumulative excess monthly mortality rate over January 2020–December 2021 for six European countries. Right: Bivariate plots of pairs of excess rates (lower triangular), 1-dimensional summaries for individual countries (diagonal), and probabilities that the excess for the country labelled on the left exceeds the rate for the country labelled at the top (upper triangular). These probabilities are the fraction of points that lie below the red line in the corresponding bivariate plot.

Extended Data Fig. 4 Excess rates over time and ranking probabilities for select countries.

: Left: excess monthly mortality rate (per 100,000 of population), with 95% uncertainty intervals, by month, for six European countries. The panels to the right display the ranking probabilities for each of the individual countries.

The probabilities sum to one for each country within each month, i.e., a country has to be in one of the six ranking positions. Also for a fixed month and ranking position across all countries the probabilities sum to one because one of the countries has to be in rank 1, 2, etc, in each month.

Extended Data Table 1 Cumulative reported COVID-19 and excess mortality measures by region
Extended Data Table 2 Cumulative reported COVID-19 and excess mortality measures by income group

Supplementary information

Supplementary Information

Comparison of WHO estimates with those of *The Economist* and IHME, country rankings on excess mortality estimates, additional analyses for Germany and Sweden, estimates for India and tables of annual estimates by country (these sections include Supplementary Figs. 1–14 and Supplementary Tables 1–13).

Reporting summary

Peer Review File

Source data

Source Data Fig. 1.

Source Data Fig. 2.

Source Data Fig. 3.

Source Data Fig. 4.

Source Data Fig. 5.

Source Data Fig. 6.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Msemburi, W., Karlinsky, A., Knutson, V. *et al.* The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature* **613**, 130–137 (2023). https://doi.org/10.1038/s41586-022-05522-2

- Received: 19 May 2022
- Accepted: 03 November 2022
- Published: 14 December 2022
- Issue Date: 05 January 2023

- DOI: https://doi.org/10.1038/s41586-022-05522-2

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

COVID deaths: three times the official toll

- Nick Petrić Howe
- Benjamin Thompson

Nature Nature Podcast 14 Dec 2022

Global estimates of excess deaths from COVID-19

- Enrique Acosta

Nature News & Views 14 Dec 2022

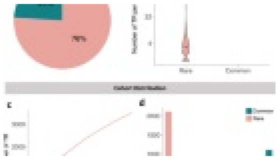
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05522-2



PanCancer analysis of somatic mutations in repetitive regions reveals recurrent mutations in snRNA U2

14 March 2022

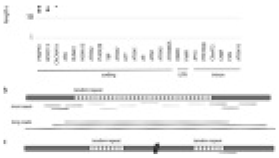
Pablo Bousquets-Muñoz, Ander Díaz-Navarro, … Xose S. Puente



Large scale in silico characterization of repeat expansion variation in human genomes

08 September 2020

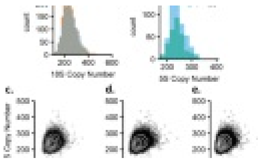
Sarah Fazal, Matt C. Danzi, … Stephan Zuchner



Long-read sequencing for rare human genetic diseases

27 September 2019

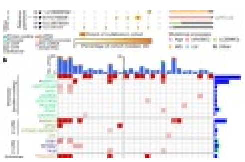
Satomi Mitsuhashi & Naomichi Matsumoto



[Thousands of high-quality sequencing samples fail to show meaningful correlation between 5S and 45S ribosomal DNA arrays in humans]

11 January 2021

Ashley N. Hall, Tychele N. Turner & Christine Queitsch



[Analyses of non-coding somatic drivers in 2,658 cancer whole genomes]

05 February 2020

Esther Rheinbay, Morten Muhlig Nielsen, … PCAWG Consortium



[Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes]

27 May 2020

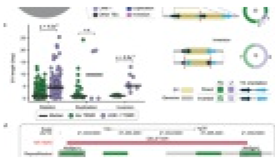
Qingbo Wang, Emma Pierce-Hoffman, … Daniel G. MacArthur



[Non-coding driver mutations in human cancer]

06 July 2021

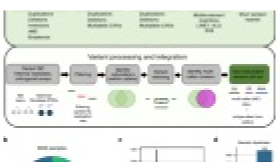
Kerryn Elliott & Erik Larsson



[Transposable element-mediated rearrangements are prevalent in human genomes]

19 November 2022

Parithi Balachandran, Isha A. Walawalkar, … Christine R. Beck



[Discovery and quality analysis of a comprehensive set of structural variants and short tandem repeats]

10 June 2020

David Jakubosky, Erin N. Smith, … Kelly A. Frazer

- Article
- Open Access
- Published: 14 December 2022

Recurrent repeat expansions in human cancer genomes

- Graham S. Erwin ,
- Gamze Gürsoy,
- Rashid Al-Abri,
- Ashwini Suriyaprakash,
- Egor Dolzhenko,
- Kevin Zhu,
- Christian R. Hoerner,
- Shannon M. White,
- Lucia Ramirez,
- Ananya Vadlakonda,
- Alekhya Vadlakonda,
- Konor von Kraut,
- Julia Park,
- Charlotte M. Brannon,
- Daniel A. Sumano,
- Raushun A. Kirtikar,
- Alicia A. Erwin,
- Thomas J. Metzner,
- Ryan K. C. Yuen,
- Alice C. Fan,
- John T. Leppert,
- Michael A. Eberle,
- Mark Gerstein &
- …
- Michael P. Snyder

Show authors

Nature **volume 613**, pages 96–102 (2023)

- 25k Accesses

- 94 Altmetric
- Metrics details

Abstract

Expansion of a single repetitive DNA sequence, termed a tandem repeat (TR), is known to cause more than 50 diseases[1,2]. However, repeat expansions are often not explored beyond neurological and neurodegenerative disorders. In some cancers, mutations accumulate in short tracts of TRs, a phenomenon termed microsatellite instability; however, larger repeat expansions have not been systematically analysed in cancer[3,4,5,6,7,8]. Here we identified TR expansions in 2,622 cancer genomes spanning 29 cancer types. In seven cancer types, we found 160 recurrent repeat expansions (rREs), most of which (155/160) were subtype specific. We found that rREs were non-uniformly distributed in the genome with enrichment near candidate *cis*-regulatory elements, suggesting a potential role in gene regulation. One rRE, a GAAA-repeat expansion, located near a regulatory element in the first intron of *UGT2B7* was detected in 34% of renal cell carcinoma samples and was validated by long-read DNA sequencing. Moreover, in preliminary experiments, treating cells that harbour this rRE with a GAAA-targeting molecule led to a dose-dependent decrease in cell proliferation. Overall, our results suggest that rREs may be an important but unexplored source of genetic variation in human cancer, and we provide a comprehensive catalogue for further study.

Main

Expansions of tandem DNA repeats (TRs) are known to cause more than 50 devastating human diseases, including Huntington's disease and fragile X syndrome[1,2]. TR tracts that cause human disease are typically large (more than 100 bp)[1]. However, identifying large TRs with short-read DNA sequencing methods is difficult because the repeat sequences are ubiquitous in the genome and many are too large—larger than the typical sequencing read length—to uniquely map to the reference genome[9]. Thus, many large TRs go undetected with current genomic technologies, and, despite their

importance to monogenic disease, the frequency and function of recurrent repeat expansions (rREs) are unknown in complex human genetic diseases such as cancer[10].

Previous studies have profiled the landscape of alterations in short TRs (STRs) in cancer genomes[3,4,5]. In particular, microsatellite instability (MSI)[6,7,8], defined by an alteration in the lengths of STRs, is prevalent in various types of cancer, including in endometrial (30%), stomach (20%) and colorectal (15%) cancers[3,4,11,12,13]. However, systematic analysis of the frequency of genome-wide large TR expansions has not been studied in cancer even though such expansions were posited to exist more than 25 years ago[14].

Recently, new bioinformatic tools to identify repeat expansions in short-read whole-genome sequencing (WGS) datasets[15,16,17,18] have led to the identification of both known and novel repeat expansions in human disease, primarily in the area of neurological disorders where repeat expansions have historically been studied[15,16,17,18,19,20,21,22,23]. Here we analysed 2,622 human cancer genomes with matching normal samples for the presence of somatic repeat expansions. We identified 160 recurrent repeat expansions (rREs) in seven types of cancer, including many rREs located in or near known regulatory elements. One of these rREs was observed in 34% of kidney cancers, and targeting this repeat expansion with sequence-specific DNA binders led to a dose-dependent decrease in cellular proliferation. Overall, our approach identifies a new class of recurrent changes in cancer genomes and provides an initial resource of these changes.

Recurrent repeat expansions

We collected uniformly processed alignments of WGS data for tumour–normal pairs in the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA), both a part of the Pan-Cancer Analysis of Whole Genomes (PCAWG) datasets[24]. After filtering, these data consisted of 2,622 cancer genomes from 2,509 patients across 29 different cancer types (Extended Data Fig. 1). Each cancer type was treated as its own cohort and was analysed independently of the other cancer types. We called somatic

rREs with ExpansionHunter Denovo (EHdn) (Methods), which measures TRs whose length exceeds the sequencing read length in short-read sequencing datasets[25,26]. That is, EHdn performs case–control comparisons using a non-parametric statistical test to determine whether repeat length is longer in tumour genomes than in matching normal genomes. This approach is analogous to joint population-level genotyping.

We first confirmed the accuracy of EHdn by performing whole-genome short- and long-read sequencing on the 786-O and Caki-1 cancer cell lines. We found that EHdn captured 72% of the repeat expansions observed in long-read sequencing (Extended Data Fig. 2). We also tested the effect of sequencing coverage on the detection of rREs and found that EHdn was robust down to 30× coverage (Extended Data Fig. 2). We then analysed 2,622 matching tumour and normal genomes with EHdn (285,363 TRs). We identified 578 candidate rREs (locus-level false discovery rate (FDR) < 10%).

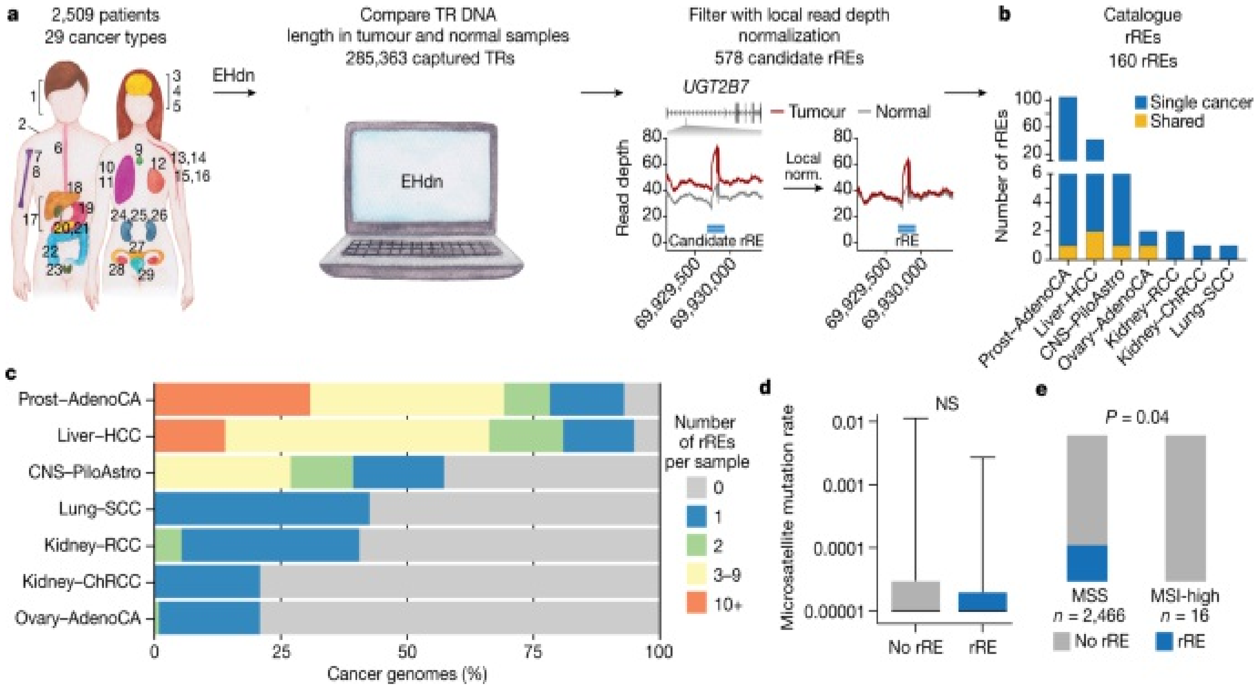
EHdn is expected to be sensitive to the copy number variations observed in cancer genomes. To account for copy number variants, we devised and implemented a local read depth filtering method that normalizes the signal originating from repeat reads using the read depth in the vicinity of the TR (Methods and Extended Data Fig. 3). We benchmarked the local read depth normalization approach with simulated chromosomal amplifications ranging from two (diploid) to ten copies. We found that this filter accounted for changes in chromosomal copy number in a manner superior to standard global read depth normalization (Extended Data Fig. 3). Overall, we conclude that local read depth normalization is valuable to identify bona fide rREs in cancer genomes and that many of the rREs that pass the filter are expanded in cancer. For example, without local read depth normalization, we could detect only 31% of candidate rREs in independent cohorts of matching tumour–normal tissue samples for breast, prostate and kidney cancers (15, 18 and 12 patients, respectively). Our local read depth filtering approach removed more than 75% (418/578) of false-positive candidate rREs (Extended Data Fig. 3). Notably, several rRE candidates that were removed are situated in hotspots for chromosomal amplification, such as chromosomal 8q amplifications that increase *MYC* production in breast cancer (Extended Data Fig. 3)[27]. Our analysis suggests that the standalone EHdn method may have selected these loci owing to amplification rather than repeat expansions, and their removal

is thus important.

After implementing our local read depth filtering strategy, we increased our detection rate to 57% (8/14) in independent cohorts (Extended Data Fig. 3). Notably, the loci we could not validate had lower expansion frequencies (5–12%). These rREs may be real but may also have been more difficult to validate in the small validation cohorts (Supplementary Table 6). Thus, we believe that this number may be an underestimate of the independent detection rate. Of the 14 candidate rREs that failed our local read depth filter, 29% (4/14) were detected in independent cohorts of samples, indicating that the filtering removes most loci that cannot be validated (Extended Data Fig. 3), but removes some true positives as well.

After accounting for local read depth, we detected 160 rREs in seven human cancer types (rRE catalogue v1.0; Fig. 1). We expected high concordance with ExpansionHunter given that this tool is related to EHdn, and indeed we observed 91% concordance with ExpansionHunter (Extended Data Fig. 4). We found that most (80%) of these loci were rarely expanded in the general population (<5% of the time, $n = 6{,}514$ genomes; Extended Data Fig. 2). rREs were primarily observed in prostate and liver cancers, but we also detected rREs in ovarian, pilocytic astrocytoma, renal cell carcinoma (RCC), chromophobe RCC and squamous cell lung carcinoma. Thus, rREs are found in tissues derived from each of the three primary germ layers (ectoderm, mesoderm and endoderm), suggesting that these expansions are a phenomenon inherent to the human genome rather than any tissue-specific process. We next performed a preliminary analysis to estimate the presence of somatic repeat expansions in individual cancer genomes. In prostate and liver cancers, most cancer genomes (93% and 95%, respectively) contained at least one rRE, with some genomes harbouring several rREs (Fig. 1c). For some pathogenic repeats, a larger TR length at birth predisposes an individual to somatic repeat expansions later in life[1,2], but we did not generally observe this with rREs (Supplementary Table 7). Overall, rREs were found in 7 of the 29 human cancer types examined and were largely cancer subtype specific.

Fig. 1: Genome-wide detection of rREs in cancer genomes.



a, Scheme of the method to identify rREs in 2,509 patients across 29 human cancer types: 1, head and neck squamous cell carcinoma (Head−SCC); 2, skin–melanoma; 3, glioblastoma (CNS–GBM); 4, medulloblastoma (CNS −Medullo); 5, pilocytic astrocytoma (CNS–PiloAstro); 6, oesophageal adenocarcinoma (Oeso−AdenoCA); 7, osteosarcoma (Bone−Osteosarc); 8, leiomyosarcoma (Bone−Leiomyo); 9, thyroid adenocarcinoma (Thy–AdenoCA); 10, lung adenocarcinoma (Lung−AdenoCA); 11, lung squamous cell carcinoma (Lung−SCC); 12, mammary gland adenocarcinoma (Breast −AdenoCA); 13, B cell non-Hodgkin lymphoma (Lymph−BNHL); 14, chronic lymphocytic leukaemia (Lymph−CLL); 15, acute myeloid leukaemia (Myeloid−AML); 16, myeloproliferative neoplasm (Myeloid−MPN); 17, biliary adenocarcinoma (Biliary–AdenoCA); 18, hepatocellular carcinoma (Liver−HCC); 19, stomach adenocarcinoma (Stomach−AdenoCA); 20, pancreatic adenocarcinoma (Panc−AdenoCA); 21, pancreatic neuroendocrine tumour (Panc−Endocrine); 22, colorectal adenocarcinoma (ColoRect–AdenoCA); 23, prostatic adenocarcinoma (Prost−AdenoCA); 24, chromophobe renal cell carcinoma (Kidney–ChRCC); 25, renal cell carcinoma (Kidney–RCC); 26, papillary renal cell carcinoma (Kidney −pRCC); 27, uterine adenocarcinoma (Uterus−AdenoCA); 28, ovarian adenocarcinoma (Ovary−AdenoCA); 29, transitional cell carcinoma of the bladder (Bladder−TCC). **b**, Distribution of rREs across cancer types. **c**,

Proportion of cancer genomes with rREs. **d**, STR mutation rate for cancer genomes with and without an rRE. Two-tailed Mann–Whitney test (*n* = 2,465 cancer genomes); NS, not significant. Boxes extend from the 25th percentile to the 75th percentile, the centre line represents the median and whiskers represent minima and maxima. **e**, Distribution of rREs across MSS and MSI-high cancers. Chi-squared (two-tailed) test with Yates' correction (*n* = 2,482 cancer genomes).

We next examined whether rREs correlate with changes in MSI[3,4]. We determined whether samples harbouring an rRE had a higher mutation rate in STRs, which is a hallmark of MSI[3,28]. We did not observe any significant difference in STR mutation rate for genomes with an rRE compared with those lacking an rRE (two-tailed Wilcoxon rank-sum test, $P = 0.27$; Fig. 1d). We also compared cancer genomes harbouring rREs with cancer genomes previously identified as MSI, using recent results from the PCAWG consortium[28]. We did not observe any enrichment in MSI for samples harbouring an rRE and instead found a weak but significant preference for rREs in microsatellite-stable (MSS) samples, not MSI samples (two-tailed Wilcoxon rank-sum test, $P = 0.04$; Fig. 1e and Extended Data Fig. 5). Thus, our findings might suggest a model in which rREs are formed by a process that is distinct from MSI.

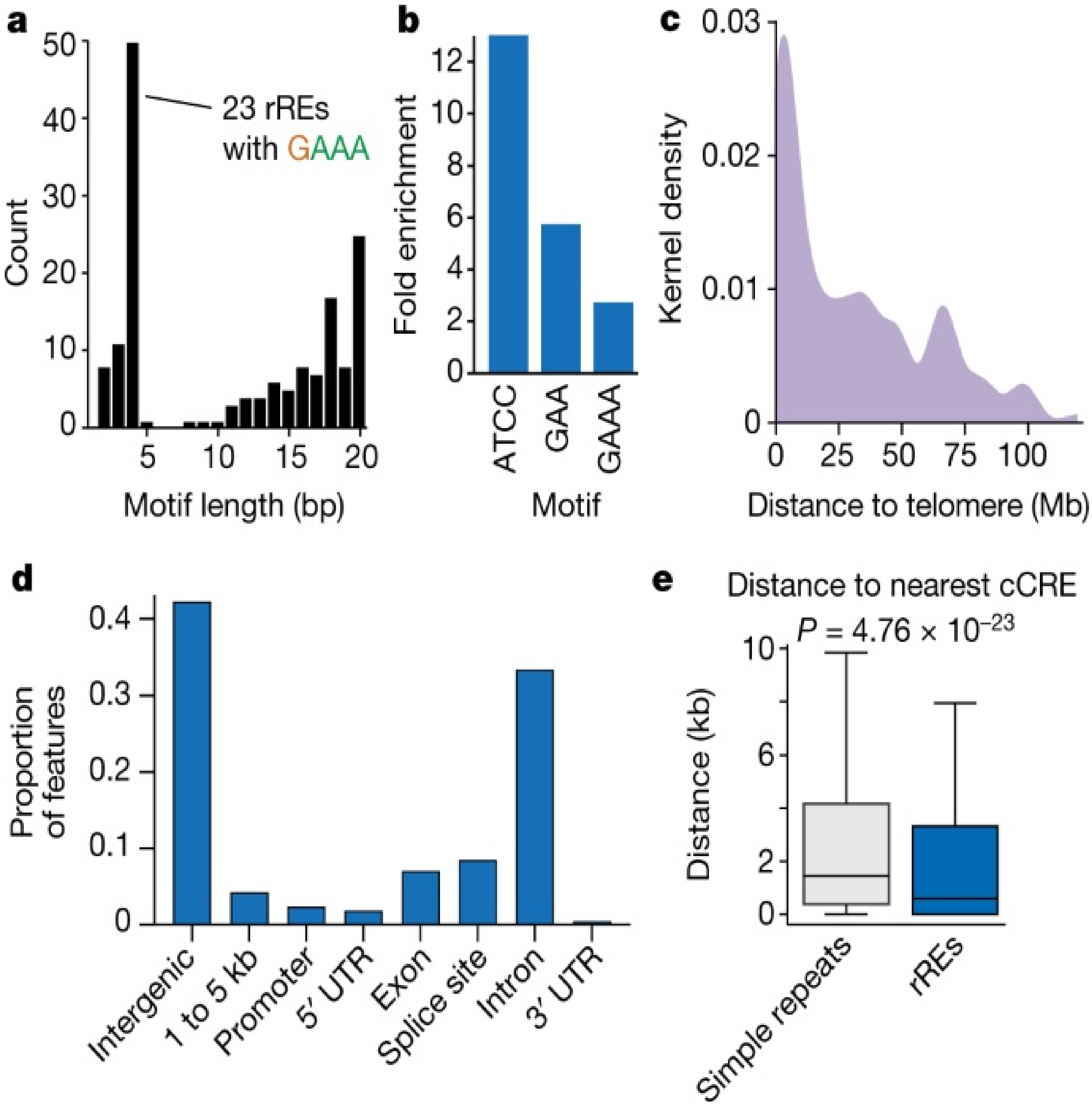
In addition to MSI, different mutational processes lead to a signature of somatic mutations. We tested whether rREs are associated with known mutational signatures by comparing them to 49 signatures of single-base substitutions (SBS) and 11 signatures of doublet-base substitutions (DBS)[29]. We performed multiple linear regression to predict the number of rREs in a sample on the basis of SBS and DBS signatures. Only one DBS signature, DBS2, showed a very weak association with rREs ($r^2 = 0.12$) (Extended Data Fig. 5).

Some rREs overlap regulatory elements

Among the 160 rREs, we observed a variety of different motifs (Supplementary Table 1) whose repeat unit length followed a bimodal distribution, in line with REs identified in other diseases (Fig. 2a and Extended Data Figs. 6 and 7)[26]. rREs were distributed across a range of G+C

content, and approximately half (76/160) had a G+C content of less than 50% (Supplementary Table 1). Six rREs contained a known pathogenic motif, all of which were GAA[30]. We examined whether any motifs were enriched in the rRE catalogue as compared with the Tandem Repeat Finder (TRF) catalogue. Although this enrichment could arise from a biological and/or technical process, we found that one of the three enriched motifs was GAA (Fig. 2b). As an example, Friedreich's ataxia is caused by a repeat expansion of a GAA motif in the intron of the gene encoding frataxin. This expansion results in DNA methylation and the deposition of repressive chromatin marks, leading to robust repression of the gene and development of disease[30]. Because of this, we suspect that some of the rREs found in cancer might alter the epigenome and affect gene regulatory networks.

Fig. 2: Features of rREs.



a, Distribution of the repeat unit (motif) for rREs. **b**, Motifs enriched in the catalogue of rREs. **c**, Distance of rREs to the end of the chromosome arm. **d**, Proportion of genic features that overlap with rREs. **e**, Distance of simple repeats ($n = 950{,}091$ loci) and rREs ($n = 160$ loci) to the nearest Encyclopedia of DNA Elements (ENCODE) cCRE. Centre values represent the median. Welch's t test (two tailed).

rREs were distributed non-uniformly across the genome, with a bias towards the ends of chromosome arms (Fig. 2c and Extended Data Fig. 6). This

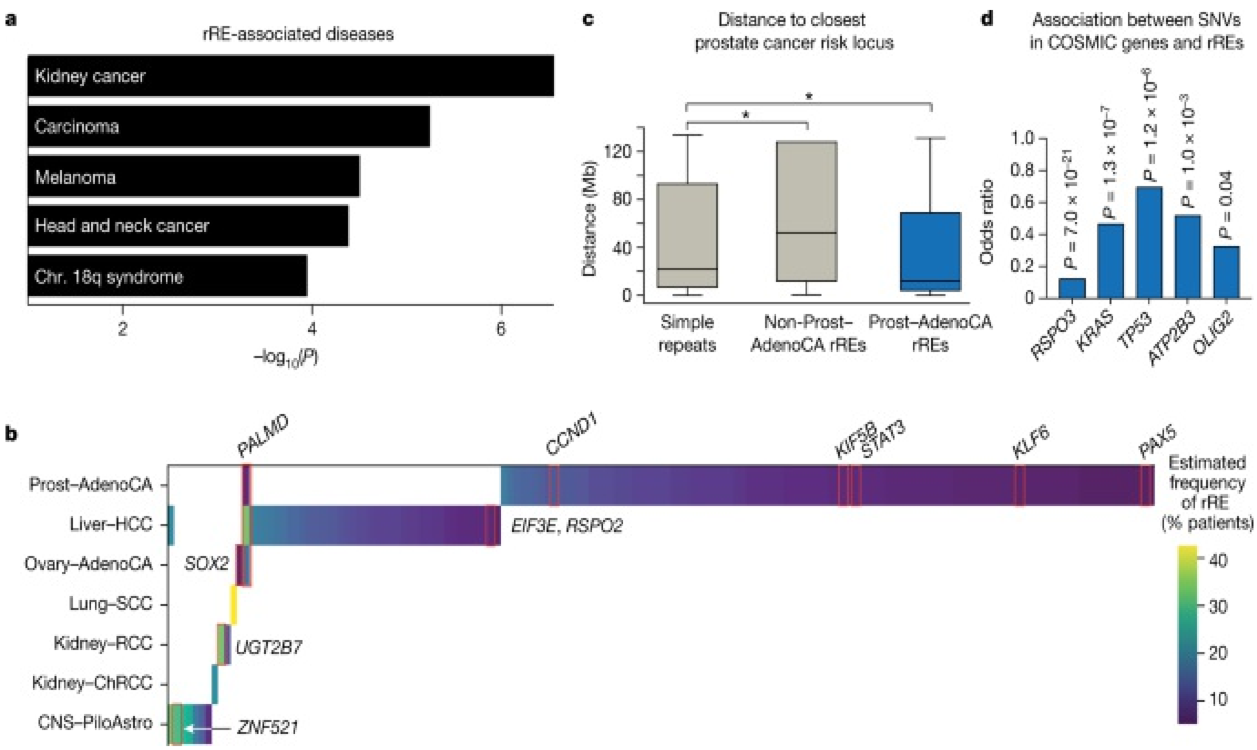
observation is consistent with previous reports of TRs and structural variants[16,31]. We also examined the distribution of rREs relative to gene features with annotatr (Fig. 2d)[32]. The 7% of rREs labelled as exonic appeared proximal to, but not within, exons, but others were in introns, untranslated regions (UTRs) and splice sites. These results suggest that rREs may have different functional roles in the regulation of gene expression.

We measured the distance between rREs and candidate *cis*-regulatory elements (cCREs)[33]; cCREs comprise approximately 1 million functional elements, including promoters, enhancers, DNase-accessible regions and insulators bound by CCCTC-binding factor (CTCF). An rRE near a regulatory element could alter the function of that regulatory element, as is observed in fragile X syndrome and Friedreich's ataxia[1]. Interestingly, rREs were located closer to cCREs than expected by chance, and we found that 54 of the 160 rREs directly overlapped with a known cCRE (Welch's t test, $P = 4.76 \times 10^{-23}$; Fig. 2e and Extended Data Fig. 7). Thus, rREs are often found in or near functional regions of the genome.

rREs with a connection to cancer

We mapped each rRE to the nearest gene and found that nine rREs mapped to tier 1 genes present in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (Fig. 3 and Supplementary Table 1). We also observed a strong correlation with cancer-related genes (Jensen disease–gene associations[34]). That is, four of the top five diseases associated with the collection of 160 rREs were cancers (Fig. 3b and Supplementary Table 4).

Fig. 3: Association of rREs with cancer.



a, Association of rREs with human diseases. Chr., chromosome. **b**, Estimated frequency of rREs in genes of interest, including nine COSMIC genes. **c**, Distance of simple repeats (n = 950,091 loci), non-prostate cancer rREs (n = 55 loci) and prostate cancer rREs (n = 105 loci) to the nearest prostate cancer risk locus. Centre values represent the median. Statistical significance was measured with Welch's t test (two tailed; *, $q = 0.08$). See Methods section 'Statistics and reproducibility' for more information. **d**, Association between SNVs in genes in the COSMIC tier 1 genes and the presence of rREs. Two-tailed Student's t test with FDR correction by the Benjamini–Hochberg method.

To examine whether some rREs have a role in oncogenesis, we looked at their association with previously identified cancer risk loci. Many rREs were identified in prostate cancer, and 63 loci have previously been associated with susceptibility to prostate cancer from available genome-wide association studies[35]. When we examined the colocalization of rREs and cancer risk loci in prostate cancer, we found that rREs were located closer to prostate cancer susceptibility loci than standard STRs or than would be expected by chance (Student's t test, FDR $q = 0.08$; Fig. 3c and Extended Data Fig. 7).

We next studied the relationship between the occurrence of COSMIC genes and the occurrence of rREs (Fig. 3d). Interestingly, after correcting for multiple-hypothesis testing, somatic mutations were found to occur significantly more in patients' genomes without rREs for five COSMIC genes. Among these genes, *TP53* in particular is striking, as wild-type *TP53* is critical for mediating the pathogenic effects of repeat expansions in both amyotrophic lateral sclerosis (ALS) and Huntington's disease[36,37]. In line with these findings, the product of the *RAD53* DNA damage repair gene in yeast is phosphorylated and activated in the presence of an expanded repeat[38].

MSI-high cancers are often correlated with higher levels of immune cell infiltration[39]. We considered whether some rREs might also be associated with higher immune cell infiltration, but we did not observe a correlation between cytotoxic activity[40] and the presence of an rRE (Extended Data Fig. 8). Because there were matching RNA sequencing (RNA-seq) data for only 4 of the 160 rREs, this analysis warrants further investigation as more matching WGS and RNA-seq datasets become available.

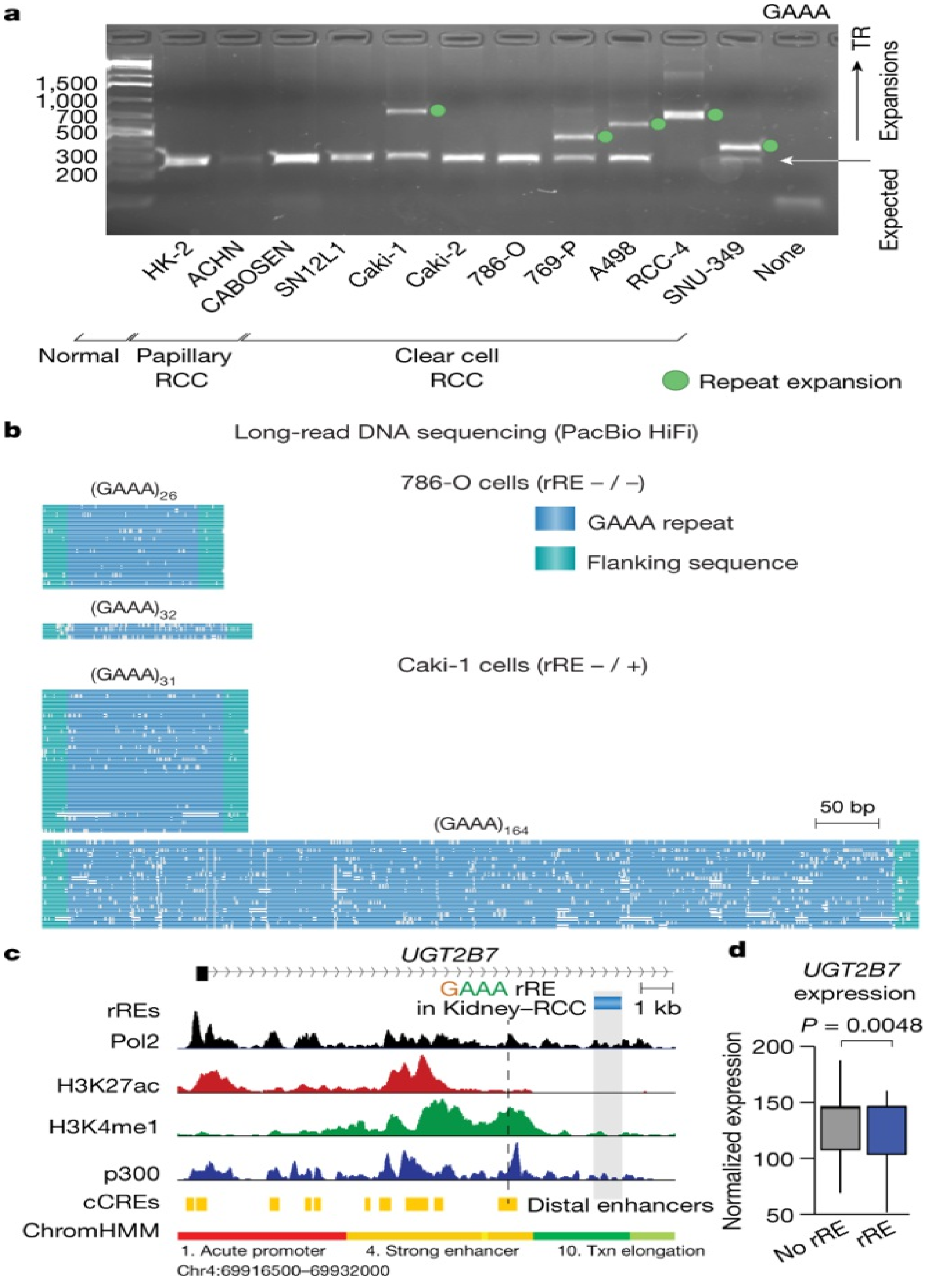
An intronic rRE detected in RCC

A GAAA expansion located in the intron of *UGT2B7* was observed in 34% of RCC samples. *UGT2B7* encodes a glucuronidase that clears small molecules—including chemotherapeutics—from the body and is selectively expressed in the kidney and liver[41].

With gel electrophoresis, we identified the expected TR size of ~26 GAAA repeats in the normal kidney cell line HK-2, corresponding closely to the length observed in the reference genome (Fig. 4a). By contrast, we identified an expansion of between ~63 and ~160 GAAA repeat units in five of eight clear cell RCC cell lines. Most expansions were heterozygous (Fig. 4a). Long-read DNA sequencing with highly accurate PacBio HiFi reads confirmed the PCR results and showed the precise structure of this repeat expansion at single-base-pair resolution for both the 786-O and Caki-1 cell lines (Fig. 4b). We also detected this repeat expansion in 5 of 12 primary kidney tumour tissue samples from patients with clear cell RCC (Extended Data Fig. 9), which showed more heterogeneity than the RCC cell lines; more

heterogeneity for human tumour samples than for clonal cell lines might be expected.

Fig. 4: An rRE in RCC.


a
GAAA
1,500
1,000
700
500
300
200
TR
Expansions
Expected
HK-2
ACHN
CABOSEN
SN12L1
Caki-1
Caki-2
786-O
769-P
A498
RCC-4
SNU-349
None
Normal
Papillary RCC
Clear cell RCC
Repeat expansion
b
Long-read DNA sequencing (PacBio HiFi)
786-O cells (rRE – / –)
(GAAA)26
(GAAA)32
GAAA repeat
Flanking sequence
Caki-1 cells (rRE – / +)
(GAAA)31
(GAAA)164
50 bp
c
UGT2B7
GAAA rRE
in Kidney–RCC
1 kb
rREs
Pol2
H3K27ac
H3K4me1
p300
cCREs
Distal enhancers
ChromHMM
1. Acute promoter
4. Strong enhancer
10. Txn elongation
Chr4:69916500–69932000
d
UGT2B7
expression
P = 0.0048
Normalized expression
200
150
100
50
No rRE
rRE

a, Gel electrophoresis of the GAAA TR in RCC samples. This analysis was performed in duplicate, and the gel is representative of the results. The units for the ladder are base pairs. For gel source data, see Supplementary Fig. 1. **b**, Visualization of long-read sequencing of the GAAA rRE in the intron of *UGT2B7*. Data are from PacBio HiFi sequencing. **c**, The locus surrounding the rRE detected in the intron of *UGT2B7*. Signal traces of RNA polymerase II (Pol2), acetylated histone H3 lysine 27 (H3K27ac), monomethylated histone H3 lysine 4 (H3K4me1) and p300 in HepG2 cells are shown. cCREs and chromatin states (ChromHMM) are also depicted. Txn, transcription. **d**, Expression of *UGT2B7* isoform ENST00000508661.1 in RCC samples as a function of detection of the rRE in *UGT2B7* (normalized expression, counts). Centre values represent the median. Significance was measured by two-tailed Wald test with FDR correction (Benjamini–Hochberg) (n = 49 cancer genomes with matching WGS and RNA-seq data).

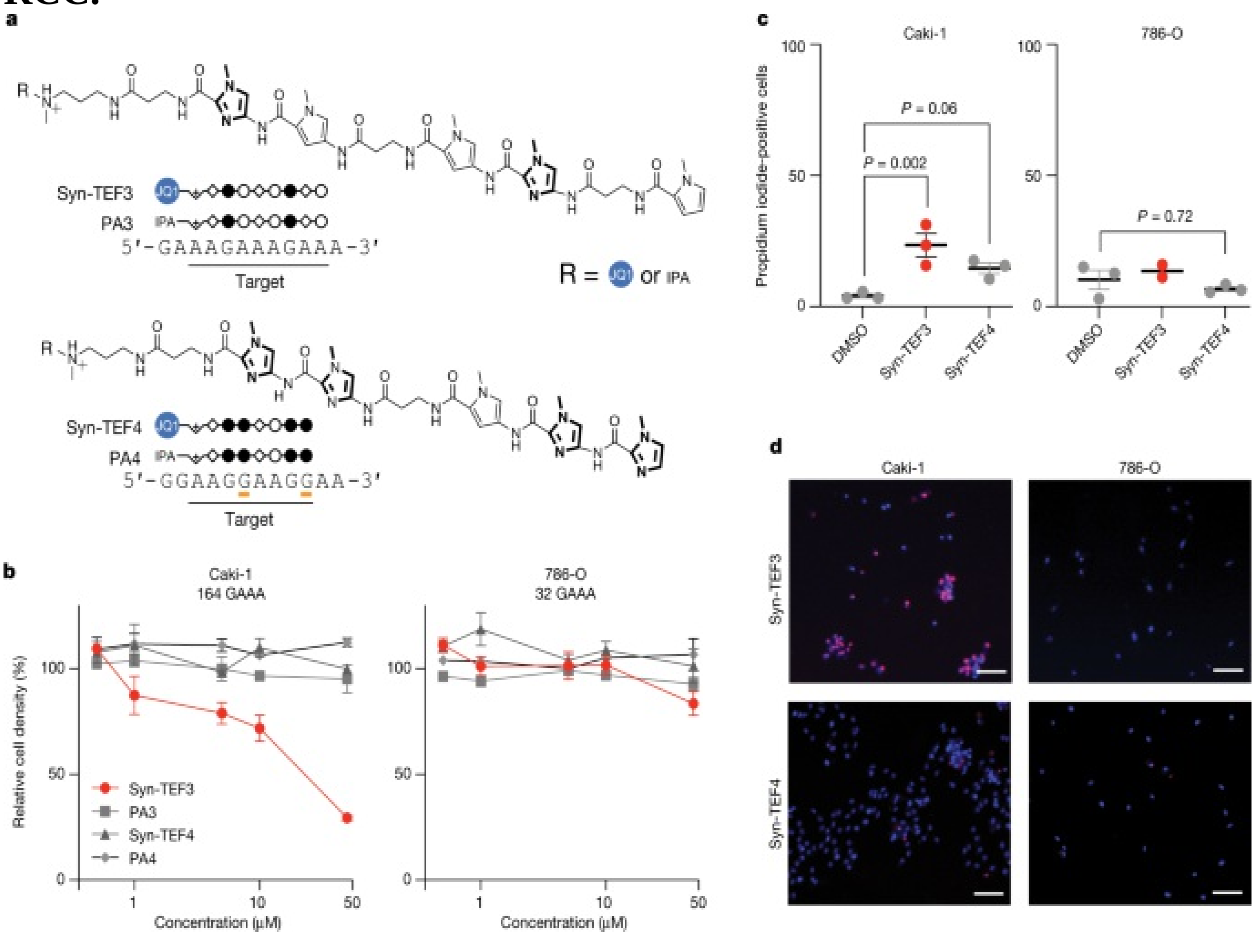
Given that *UGT2B7* is selectively expressed in the liver and kidney, and that it has a role in clearing small molecules from the body, we examined whether this rRE may be located near any functional elements that could regulate its expression. Analysis of the chromatin environment surrounding the rRE in *UGT2B7* identified a nearby enhancer, raising the possibility that this rRE alters the expression of *UGT2B7* (Fig. 4c). The repeat motif of this rRE, GAAA, appears similar to the pathogenic repeat motif found in Friedreich's ataxia, which is GAA. The pathogenic GAA-repeat expansion blocks *FXN* expression[30]. We therefore considered whether the intronic GAAA-repeat expansion might repress the expression of *UGT2B*; we found a modest decrease in expression that was not statistically significant (Extended Data Fig. 8). While this rRE was also not associated with a difference in survival (Extended Data Fig. 8), it was associated with a significant decrease in a transcript isoform of *UGT2B7* (Wald test with FDR correction, P = 0.0048) (Fig. 4e). Interestingly, a shift in isoform usage of *UGT2B7* has been noted in cancer[42].

Repeat-targeting molecules

Do GAAA-repeat expansions contribute to cell proliferation? Targeting pathogenic repeat expansions with small molecules has been demonstrated

previously[43]. We previously showed that targeting a related TR motif, GAA, with synthetic transcription elongation factors (Syn-TEF1) reverses pathogenesis in several models of Friedreich's ataxia[44]. Therefore, if the GAAA rRE in RCC behaves similarly, a Syn-TEF targeting GAAA might have anti-proliferative activity. We rationally designed Syn-TEF3, which contains a GAAA-targeting polyamide and a bromodomain ligand, JQ1, designed to recruit part of the transcriptional machinery (Fig. 5a and Supplementary Fig. 2). We also included a control molecule, Syn-TEF4, which targets GGAA TRs, as well as polyamides PA3 and PA4 that lack the JQ1 domain. We have previously shown that Syn-TEFs and polyamides localize to repetitive TRs in living cells[44,45].

Fig. 5: Design and characterization of GAAA-targeting molecules in RCC.



a, Chemical structures of Syn-TEF3, PA3, Syn-TEF4 and PA4. Syn-TEF3 and PA3 target 5′-AAGAAAGAA-3′. Syn-TEF4 and PA4 target 5′-

AAGGAAGG-3′. The structures of *N*-methylpyrrole (open circles), *N*-methylimidazole (filled circles) and β-alanine (diamonds) are shown. *N*-methylimidazole is bolded for clarity. The structure of JQ1 linked to polyethylene glycol (PEG_6) is represented as a blue circle. The structure of isophthalic acid and its linker is represented as IPA. Complete chemical structures appear in Supplementary Fig. 2. Mismatches formed with Syn-TEF4 and PA4 are indicated with orange lines. **b**, Relative cell density of RCC cell lines Caki-1 and 786-O following treatment (72 h) with compounds as indicated. Relative cell density was measured by CCK-8 assay (Methods). Results are shown as the mean ± s.e.m. ($n = 4$ biological replicates). **c**, Quantification of the percentage of propidium iodide-positive cells. *P* values are from one-way ANOVA with Bonferroni's correction for multiple comparisons. Results are shown as the mean ± s.e.m. ($n = 3$ biological replicates except $n = 2$ biological replicates for Syn-TEF3 in 786-O cells). **d**, Live-cell microscopy of Caki-1 and 786-O cells stained with propidium iodide (red) and Hoechst 33342 (blue). Scale bars, 100 μm. See also Extended Data Fig. 10.

We examined the effect of Syn-TEFs on cell proliferation (Fig. 5b). Caki-1 and 786-O cells were selected because they have the largest (164 repeats) and smallest (32 repeats) GAAA tracts, respectively, within the first intron of *UGT2B7*. We observed that Syn-TEF3 led to a significant decrease in the proliferation of Caki-1 cells in a dose-dependent manner, but had little effect on 786-O cells. Syn-TEF4, which does not target GAAA TRs, did not significantly decrease proliferation in either of the cell lines tested, demonstrating a requirement for GAAA-specific targeting (Fig. 5b). Two additional cell lines with GAAA-repeat expansions as well as two additional control non-expanded cell lines showed a similar association between Syn-TEF sensitivity and presence of the repeat expansion (Extended Data Fig. 10). In line with this finding, Caki-1 cells treated with Syn-TEF3 exhibited a significant increase in cell death when compared with the DMSO-treated control, as measured by propidium iodide staining (Fig. 5c,d and Extended Data Fig. 10). By contrast, 786-O cells treated with Syn-TEF3 showed no significant difference in propidium iodide-positive cells when compared with DMSO-treated cells (Fig. 5c,d and Extended Data Fig. 10). Notably, the Syn-TEF4, PA3 and PA4 control agents had no significant effect on cell death in either cell line when compared with vehicle control (Fig. 5c,d and Extended

Data Fig. 10). These results are preliminary and warrant further study, but they suggest that GAAA-repeat expansions may represent a genetic vulnerability in RCC.

Discussion

Here we conducted a genome-wide survey of rREs, distinct from MSI, across cancer genomes. Our data (1) identified 160 rREs in seven human cancer types and showed that (2) most (155 of 160) rREs are cancer subtype specific; (3) among diseases, rREs are enriched in human cancer loci and tend to occur near regulatory elements; (4) rREs do not correlate with MSI status; and (5) targeting a GAAA-repeat expansion in RCC with a small molecule leads to cancer cell killing. Taken together, our results uncover an unexplored genetic alteration in cancer genomes with important mechanistic and therapeutic implications.

Cancer cells evolve and adapt in response to environmental or pharmacological perturbations, but the mechanisms supporting these changes are still being uncovered. One source of genetic variation that may enable genetic adaptations is TR DNA sequences. Mutations in the repeat length of TRs can occur up to 10,000 times more frequently than single-nucleotide variants (SNVs) or insertions and deletions (indels)[1]. Repeat expansions may provide a source of genetic variation to enable cancer cells to adapt to changes in the environment[46]. Indeed, colorectal cancers acquire mutations in STRs in response to targeted therapy just 24 h after treatment, suggesting that mutations in these regions may associate with rapid evolution[47]. In future studies, it will be particularly valuable to study repeat expansions in the genomes of cancer cells that face changing environments, including metastasis and chemotherapy.

Historically, MSI has been the focus of efforts to profile changes in STRs in cancer genomes because specific cancer-causing genetic alterations in repair genes can promote widespread STR alterations. Interestingly, we find little to no correlation between rREs and MSI. These results are consistent with previous findings in which the correlation between MSI and repeat instability at larger TRs was not definitive[48]. MSI may contribute to a subtype of rREs that we have not yet uncovered, or rREs may arise from a mutation process

that is distinct from that of MSI. There are several different cellular repair systems for DNA, and the rREs we observed are presumably due to very specific locus-associated mechanisms or activities. Some of these repeat expansions may be due to *cis* regions with interesting DNA or chromatin configurations that are prone to expansion at distinct loci, rather than gene mutations that cause global *trans* effects, as occurs in MSI.

There are numerous mechanisms by which a repeat expansion can alter cellular function. Known pathogenic repeat expansions can alter the coding sequence of a protein, such as in the case of Huntington's disease. However, several repeat expansions in non-coding regions alter gene expression[1]. In other instances, the repeat expansion can lead to a pathogenic RNA molecule (myotonic dystrophy) or protein (ALS)[1]. Finally, repeat expansions in MSI-associated cancers, which are too small to detect by EHdn, can disrupt DNA replication[49]. Thus, our catalogue represents a powerful resource to explore the mechanisms by which rREs alter cellular function in cancer.

Tools to identify repeat expansions are still in their infancy. The field would benefit from cohorts of samples with whole-genome long-read DNA sequencing data, improved bioinformatic methods, increased sequencing coverage and increased cohort sizes. As with other tools that identify repeat expansions from short-read sequencing data, we cannot distinguish zygosity from sample heterogeneity or obtain the precise lengths of repeats. Our independent experimental validation showed that some repeat expansions are heterogeneous (Extended Data Fig. 8). We suspect that tumour heterogeneity may lead to an under-reporting of rREs. Furthermore, this study focuses on somatic mutations, but repeat expansions that occur in the context of normal development will be another important area of study[10]. Furthermore, germline events that predispose an individual to cancer would also be worth studying; there is evidence that a TR in the androgen receptor gene is associated with tumour stage and tumour grade at prostate cancer onset[50]. Finally, we only detected changes in repeat length that were greater than sequencing read length. In future studies, it will be important to explore recurrent changes that are smaller in length. Finally, it is important to acknowledge that rREs could be mediators of phenotypes or passengers that result from genetic instability and clonal selection. In the one instance where we targeted an rRE in RCC, cell proliferation was reduced, in line with a

mediator role for this rRE. Distinguishing between these two possibilities for each rRE will be an important line of work in the future.

To our knowledge, this is the first genome-wide survey of repeat expansions beyond a neurological or neurodegenerative disorder. Thousands of high-quality whole-genome sequences exist for many diseases, and our data provide evidence that repeat expansions should be explored beyond the classical bounds of neurodegenerative diseases where they have been most investigated. Our results provide a framework to analyse WGS datasets from complex diseases such as cancer.

Methods

Data curation

We obtained white-listed data from the ICGC and TCGA PCAWG dataset. The term 'white-listed' refers to samples that passed quality control by the PCAWG consortium[24]. Data were accessed through the Cancer Genome Collaboratory. We used aligned reads (BAM files), which were aligned to GRCh37 as described previously[24]. These data are available through the PCAWG data portal (https://docs.icgc.org/pcawg). A list of samples included in the analysis is available in Supplementary Table 2.

Identification of somatic rREs

We analysed tumour and matching normal samples for each cancer type independently. We executed EHdn (v0.9.0)[16] with the following parameters: --min-anchor-mapq 50 --max-irr-mapq 40. To prioritize loci, we developed a workflow termed Tandem Repeat Locus Prioritization in Cancer (TROPIC). We included loci from chromosomes 1–22, X and Y for downstream analysis. We removed loci where >10% of Anchored in-repeat read (IRR) values were >40, which is the theoretical maximum value. The *P* value (from a non-parametric one-sided Wilcoxon rank-sum test) for each locus was used to calculate an FDR *q* value. Loci with FDR < 0.10 are reported. We selected loci where >5% of samples had an Anchored IRR quotient of >2.5. The

results of our filtering are available in Supplementary Table 3. For a repeat expansion to be detected by EHdn, the TR was required to be larger than the sequencing read length. A somatic repeat expansion was defined as having FDR $q < 0.05$ in a comparison of the tumour and normal samples. We next calculated a preliminary estimate of the frequency of rREs in each cancer. To call repeat expansions in individual cancer samples, we analysed the distribution of tumour and normal Anchored IRR values and selected a conservative threshold for the Anchored IRR quotient ((tumour Anchored IRR – normal Anchored IRR)/(normal Anchored IRR + 1)) > 2.5 (Extended Data Fig. 4).

Local read depth normalization

EHdn normalizes the number of Anchored IRRs for a given locus to the global read depth. To account for chromosomal amplifications and other forms of genetic variation that could alter local read depth, we performed the following normalization. For each rRE locus and sample in its corresponding cancer, samtools v1.13 was used with the parameter depth -r to find the read depth at each base pair within the locus and a 500-bp region encompassing the start and stop positions of the TR. We calculated the average read depth at each base pair and defined this as the local read depth. Finally, we calculated the local read depth-normalized Anchored IRR value specific to a sample and rRE combination by dividing the non-normalized Anchored IRR value from EHdn by the local read depth at the locus.

Generation of CABOSEN cells

CABOSEN cells were generated from a cabozantinib-sensitive (CABOSEN) human papillary RCC xenograft tumour grown in *Rag2*$^{-/-}$*γC*$^{-/-}$ mice, as described previously[51]. Tumour tissue was minced with a sterile blade, and the cell suspension was cultured in DMEM/F-12 medium (Corning) supplemented with 10% (vol/vol) Cosmic calf serum (ThermoFisher). Cells were expanded and cryopreserved in growth medium supplemented with 10% (vol/vol) DMSO, and cells from passage 8 were used for analysis.

Analysis of rREs by gel electrophoresis

We performed PCR with CloneAmp HiFi PCR Mix (Takara Biosciences) and added DMSO to a final concentration of 5–10% (vol/vol) as needed. A list of the primers used to analyse the loci is available in Supplementary Table 5. All cell lines tested negative for mycoplasma contamination with the MycoAlert Mycoplasma Detection kit (Lonza). Cell line identities were authenticated through STR profiling by the Genetic Resources Core Facility at Johns Hopkins University, with the exception of SNU-349 cells, which did not match the reported STR profile of SNU-349 cells or any other catalogued cell line but had a mutated *VHL* gene and expressed high levels of *PAX8* and *CA9*, in line with a clear cell RCC origin.

Visualization of repeat expansions with ExpansionHunter and REViewer

To inspect the reads supporting a repeat expansion, we annotated the repeat as described on the GitHub page for ExpansionHunter. We then profiled the region with ExpansionHunter (v4.0.2) using the default settings[15]. The resulting reads were visualized with REViewer (v0.1.1) using the default settings. REViewer is available at https://github.com/Illumina/REViewer. A repeat expansion was called when the repeat tract length for one allele of the tumour sample was greater than 100 bp and exceeded the repeat tract length of both normal alleles. A locus was considered validated if at least ten cancer genomes had a repeat expansion.

Validation of rREs in independent cohorts of samples

Twelve pairs of matching normal and tumour samples from patients with clear cell RCC were obtained with the patients' informed consent ex vivo upon surgical tumour resection (Stanford institutional review board-approved protocols 26213 and 12597) and analysed. Eighteen and 15 pairs of matching normal and tumour samples for prostate and breast cancer, respectively, were obtained from the Tissue Procurement Shared Resource facility at the Stanford Cancer Institute and analysed. These samples were obtained with

patients' informed consent (Stanford institutional review board-approved protocols 11977 and 55606). Nucleic acid was isolated with either the Quick Microprep Plus kit (D7005) or the Zymo Quick Miniprep Plus kit (D7003) (Zymo Research). Gel electrophoresis was performed as described above. A locus was considered detected if a somatic repeat expansion was identified in at least one patient tumour sample compared with a matching normal sample.

Downsampling analysis

For the downsampling analysis, tumour genomes from RCC samples were downsampled from their mean (52×) sequencing depth to 40×, 30×, 20× and 10× depth with the samtools view command. EHdn was run, as described above, for each of the sequencing depths, and the Bonferroni-corrected *P* value was plotted for the rRE in *UGT2B7* (GAAA, chr4:69929297–69930148).

Benchmarking the local read depth normalization filter

We benchmarked the local read depth filter in silico by observing its behaviour with simulated reads. First, we created a reference genome containing artificially expanded repeats. We randomly selected ten TRs located on chromosome 1 that were shorter than the sequencing read length of 100 bp. We artificially expanded these TRs on chromosome 1 of GRCh37 with the BioPython Python package (v1.79). Next, we used wgsim (v0.3.1-r13) to simulate reads from the reference file with the command 'wgsim -N 291269925 --1 100 --2 100 reference_file.fasta output.read1.fastq output.read2.fastq'. The number of reads (specified by the -N option) was calculated to achieve 30× coverage of chromosome 1. The resulting pair of files, hereafter referred to as the base fastq files, contained a copy number of 2 for all of the expansions.

To simulate copy number amplification, the read simulation process was repeated using reference files that contained only the artificially expanded repeats and their surrounding 1,000-bp flanking regions. We created ten pairs of fastq files, each with an increasing copy number. We specified the copy number by multiplying the number of reads to generate (wgsim -N option) by

the required number. To generate the final set of fastq files, we concatenated each pair of copy number-amplified fastq files with the base fastq files. The end result was eight pairs of fastq files that contained reads for chromosome 1 and copy number amplification varying from 2 to 10 of the expanded repeats.

The base fastq file with a copy number of 2, in addition to the eight copy number-amplified fastq files, was aligned to chromosome 1 of GRCh37 with bwa-mem (v0.6) with the default options. The resulting SAM files were converted to BAM format with samtools (v1.15) using the default options. Finally, we ran the EHdn profile command (v0.9.0) with the minimum anchor mapping quality set to 50 and maximum IRR mapping quality set to 40. Finally, the Anchored IRR values were extracted by overlapping the STR coordinates with the de novo repeat expansion calls.

Short-read and long-read DNA sequencing

We sequenced the Caki-1 and 786-O cell lines with both short-read sequencing (60× sequencing coverage, 150-bp paired-end sequencing on a NovaSeq 6000 instrument) and long-read sequencing (50× sequencing coverage, PacBio HiFi sequencing on a Sequel IIe instrument). We aligned the long reads to GRCh37 with pbmm2 (v1.7.0), using the parameters --sort --min-concordance-perc 70.0 --min-length 50. We aligned the short reads to GRCh37 with Sentieon (v202112.01) using parameters -K 10000000 -M, an implementation of BWA-MEM, and analysed the samples with EHdn, as described above. We included loci for which at least one sample had an Anchored IRR value of >0 for further analysis. Anchored IRR values >0 arise when the repeat length exceeds the sequencing read length. To benchmark EHdn against long-read sequencing data, we manually determined the TR length of a given locus in the long-read sequencing data. If the TR length in the long-read sequencing data exceeded the short-read sequencing read length of 150 bp, we considered that locus to have been confirmed.

The PacBio HiFi data were aligned to GRCh37 with pbmm2 (v1.7.0) and visualized at the *UGT2B7* locus with Tandem Repeat Genotyper (v0.2.0; https://github.com/PacificBiosciences/trgt).

Analysis of rRE loci

To determine whether rREs were associated with any human diseases, rREs were mapped to genes with GREAT (v4.0.4, default settings)[52]. The resulting genes were analysed with Enrichr using Jensen Diseases[53]. The output of this analysis is available in Supplementary Table 4. To determine whether repeat expansions were associated with MSI-high cancers, we obtained data from ref. [3]. The percentage of MSI-high cancers was obtained for colon adenocarcinoma (COAD), stomach adenocarcinoma (STAD), kidney renal cell carcinoma (KIRC), ovarian serous cystadenocarcinoma (OV), prostate adenocarcinoma (PRAD), head and neck squamous cell carcinoma (HNSC), liver hepatocellular carcinoma (LIHC), bladder urothelial carcinoma (BLCA), glioblastoma multiforme (GBM), skin cutaneous melanoma (SKCM), thyroid carcinoma (THCA) and breast invasive carcinoma (BRCA) and compared with the number of repeat expansions and the percentage of patients with at least one repeat expansion in the corresponding cancer type from the PCAWG dataset. We also overlapped cancer genomes containing rREs with the microsatellite mutation rate (data available for all but 157 PCAWG genomes analysed in this study), which we term the STR mutation rate, and MSI calls from ref. [28]. The association of rREs with STR mutation rate was assessed with the two-tailed Wilcoxon rank-sum test. The association of rREs with MSI calls was assessed by chi-squared test with Yates' correction.

To determine whether rREs were associated with known mutational signatures, we downloaded mutational signatures from the ICGC Data Coordination Center (DCC; https://dcc.icgc.org/releases/PCAWG/mutational_signatures/Signatures_in_Sa
We performed multiple linear regression for each SBS and DBS signature to identify predictors of the number of rREs present in a sample. To choose the predictors, we performed best subset selection on DBS and SBS signatures and included age as a possible confounding factor. We used statsmodels (v0.12.2) in Python and, specifically, the ordinary least-squares model found in the statsmodels.api.OLS module to estimate the coefficients of the selected predictors in their corresponding multiple linear regression model[54].

To determine whether repeat expansions were associated with a difference in cytotoxic activity, we calculated cytotoxic activity as previously described for four cancers that had matching RNA-seq and WGS data[40]. For each locus, we compared the cytolytic activity for patients with a repeat expansion to that for patients without a detected repeat expansion using a Welch's *t* test (a two-tailed test) with correction for multiple-hypothesis testing (Benjamini–Hochberg FDR $q < 0.05$). rREs were annotated with genic elements using annotatr (v1.18.1)[32].

To determine whether rREs were associated with regulatory elements, we downloaded cCREs[33] and mapped them to GRCh37 with LiftOver (UCSC) ($n = 950,091$ after removing 174 outliers)[55]. We determined the distance between rREs and cCREs with the bedtools closest command (v2.27.1)[56] and compared this distance to that for a simple repeats catalogue[57]. To compare the distance to ENCODE cCREs, a Welch's *t* test was performed.

To determine whether prostate cancer rREs were associated with prostate cancer susceptibility loci[35], we calculated the distance to three sets of loci using the 'bedtools closest' command. We calculated the distance between (1) rREs present in prostate cancer samples and prostate cancer susceptibility loci, (2) rREs not present in prostate cancer samples and prostate cancer susceptibility loci and (3) simple repeats and prostate cancer susceptibility loci. To compare the distances between these three associations, we performed a Welch's *t* test with FDR correction (Benjamini–Hochberg).

To determine whether rREs were associated with replication timing, we downloaded Repli-seq replication timing data for seven cell lines from the ENCODE website (NCI-H460, T470, A549, Caki2, G401, LNCaP and SKNMC)[58]. We selected regions for which all cell lines had concordant signals for analysis (early or late replication designations in agreement for each cell line at a given locus). We determined whether there was a difference in the distribution of rREs across early- and late-replicating regions compared with the simple repeats catalogue by using bootstrapping ($n = 10,000$). We sampled 54 loci (the number of rREs present in a concordant replication region) from rREs and simple repeats. A Welch's *t* test was performed on the bootstrapped samples to estimate a *P* value. We applied FDR correction (Benjamini–Hochberg) to the estimated *P* values. To

determine whether rRE status in *UGT2B7* was associated with survival outcome in patients with clear cell RCC (TCGA abbreviation, KIRC), we used Welch's *t*-test quartile.

To identify motifs enriched and depleted in the rRE catalogue, we followed the same method as in the motifscan Python module (v1.3.0)[59]. We compared our rRE catalogue to the simple repeats catalogue (TRF) as a control. For each unique motif present, we built a contingency table specifying the count of rREs and simple repeats with and without the motif. Two one-tailed Fisher's exact tests were applied to the table to test for significance in both directions, that is, enrichment and depletion. The 'stats' module in the Scipy Python package (v1.7.0) was used to conduct the significance test. Because multiple-hypothesis tests were performed, we applied FDR correction (Benjamini–Hochberg) for multiple-hypothesis testing to the *P* values, with a cut-off (FDR) of 0.01.

For the comparison of SNVs in COSMIC genes to rREs, we first divided the cancer genomes into two categories: an rRE cohort and a non-rRE cohort. The rRE cohort contained all genomes that had at least one rRE detected (n = 615), and the non-rRE cohort contained all genomes that had no rREs detected (n = 1,897). We then looked at the number of donors in the rRE cohort that had at least one mutation in a given gene (COSMIC tier 1 genes) i and the number of donors in the non-rRE cohort that had at least one mutation in a given gene i with a contingency table. We calculated the *P* value (Fisher's exact test) for the significance of associating genes with either the rRE or non-rRE cohort. This *P*-value calculation was repeated for all COSMIC genes, using FDR at a significance level of 0.05 (Benjamini–Hochberg) to correct for multiple-hypothesis testing.

Estimation of expansions in the general population

To estimate the frequency of rREs in the general population, EHdn (v0.9.0) was run on 1000 Genomes Project samples[60] (n = 2,504) (GRCh38) and Medical Genome Reference Bank[61] samples (n = 4,010) (GRCh37 lifted over to GRCh38).

The genomic coordinates of the 160 rREs (GRCh37) were padded with 1,000

bp and translated to GRCh38 coordinates with UCSC LiftOver. Then, the rRE coordinates (GRCh38) were overlapped with loci from the population samples containing Anchored IRR calls. rREs that overlapped with matching motifs in the population samples were selected for further analysis. We next sought to identify expanded rREs in the population samples to quantify their prevalence. To do so, we converted their global-normalized Anchored IRR values to be comparable to ICGC values. This step was necessary because sequencing read lengths in the PCAWG dataset are generally 100 bp while the read lengths in the 1000 Genomes and Medical Genome Reference Bank datasets are 150 bp. Conversion followed the formula (Anchored IRR, 100 bp) = 0.5 + 1.5 × (Anchored IRR, 150 bp)[16]. A sample in the population samples was counted as expanded if its Anchored IRR value was greater than the 99th percentile of Anchored IRR values in the normal samples from the PCAWG dataset, a threshold that is comparable to the threshold used to call expansions in tumour samples (Extended Data Fig. 4). In future rRE catalogues, for the rare instance where the estimated frequency of repeat expansions in the population samples is higher than expected, these data could be used to further filter rREs to improve the detection of cancer-specific repeat expansions.

To compare the length of TRs in normal samples with and without a matching rRE in a tumour sample, donors in the Prost-AdenoCA and Kidney-RCC cohorts whose data are available for download through the Cancer Collaboratory were included ($n = 253$). We used ExpansionHunter (v5.0.0) with the default options to genotype prostate and kidney cancer rREs in the normal samples of the selected donors. When there were two alleles of an rRE in a sample, both alleles were included and treated as distinct data points. For each rRE, we tested whether the distribution of genotypes from donors who had an expansion in their tumour samples differed from that for donors who did not have an expansion. Student’s t test was used to compute P values with FDR correction (Benjamini–Hochberg) to adjust for multiple-hypothesis testing.

Association of rREs with gene expression

Matching RNA-seq and WGS data were available for Kidney–RCC, Ovary–AdenoCA, Panc–AdenoCA and Panc–Endocrine. RNA-seq data from these

samples were obtained from the DCC (https://dcc.icgc.org/), and values were converted to transcripts per million (TPM). Normalized gene expression (TPM) values were compared for samples with and without an rRE (Welch's *t* test, with FDR correction). For isoform analysis, normalized gene expression counts were compared for samples with and without a repeat expansion using the DESeq2 (v1.32.0) package in R (v4.0.5). We used the DESeq function to calculate the log_2-transformed fold change for three isoforms of the *UGT2B7* gene (ENST00000305231.7, ENST00000508661.1 and ENST00000502942.1) and performed a Wald test with FDR correction using the Benjamini–Hochberg procedure (q-value threshold of $q < 0.01$).

Design, synthesis and characterization of Syn-TEFs and polyamides

Syn-TEFs and polyamides were designed to target a GAAA repeat (Syn-TEF3 and PA3) or a control GGAA repeat (Syn-TEF4 and PA4). Syn-TEF3, Syn-TEF4, PA3 and PA4 were synthesized and purified to a minimum of 95% compound purity by WuXi Apptec and used without further characterization. HPLC conditions for chemical characterization were as follows: flow rate of 1.0 ml min^{-1}; solvent A: 0.1% (vol/vol) trifluoroacetic acid (TFA) in water; solvent B: 0.075% (vol/vol) TFA in acetonitrile; Gemini column: C18 5 µm 110A 150 × 4.6 mm. Full results of characterization can be found in Supplementary Fig. 2.

Treatment of RCC cell lines with Syn-TEFs

Caki-1, 786-O and Caki-2 cells were obtained from the American Type Culture Collection (ATCC) and grown in RPMI-1640 with l-glutamine (Gibco, 11875093), supplemented with 10% (vol/vol) FBS. A498 and ACHN cells were obtained from ATCC and grown in DMEM with glucose, l-glutamine and sodium pyruvate (Corning, 10-013-CV), supplemented with 10% (vol/vol) FBS. RCC-4 cells were obtained from A. Giacca (Stanford University) and grown in DMEM with glucose, l-glutamine and sodium pyruvate (Corning, 10-013-CV), supplemented with 10% (vol/vol) FBS. Cell line identities were confirmed by STR profiling (Genetic Resource Core

Facility, Johns Hopkins University) and tested negative for mycoplasma. Cells were seeded in 96-well plates on day 0. On day 1, cells were treated with the indicated molecules. Molecules were dissolved in DMSO (vehicle) and added to cells (0.1% (vol/vol) final concentration of DMSO). On day 4 (72 h later), relative metabolic activity was measured as a proxy for relative cell density, using the Cell Counting Kit (CCK-8, Dojindo Molecular Technologies) according to the manufacturer's instructions. Absorbance (450 nm) of cells treated with molecules was normalized to that for cells treated with DMSO (0.1% (vol/vol)) or with no treatment. Absorbance was measured with an Infinite M1000 microplate reader (Tecan).

For microscopy, Caki-1 and 786-O cells were plated on glass-bottom 96-well plates under standard culture conditions. One day after plating, medium containing no drug, 50 μM Syn-TEF3 or 50 μM Syn-TEF4 was added, and the cells were incubated for 72 h at 37 °C. As a control, wells that received no treatment were incubated with 70% (vol/vol) ethanol for 30 s before staining. Cells were then stained with propidium iodide, Calcein-AM and Hoechst 33342 from the Live-Dead Cell Viability Assay kit (Millipore Sigma, CBA415) according to the manufacturer's instructions and immediately imaged at ×10 magnification with a 0.17-NA CFI60 objective on a Keyence BZ-X710 microscope. Eight fields were measured for each treatment condition, and the experiment was repeated two times. Quantification was conducted using FIJI software (release 20220330-1517). For statistical analyses, one-way ANOVA adjusted with Bonferroni correction for multiple comparisons was conducted with GraphPad Prism (v9.3.1).

Statistics and reproducibility

Data are represented as the mean ± s.e.m. unless stated otherwise. All experiments were reproduced at least twice unless stated otherwise. Box plots were prepared with matplotlib (v3.4 or v3.6) as follows unless stated otherwise: the box extends from the first quartile (Q1 or 25th percentile) to the third quartile (Q3 or 75th percentile) of the data, with a line at the median. The whiskers extend from the box by 1.5 times the interquartile range (IQR). The IQR is the difference between the values at Q3 and Q1. Outliers were not plotted to improve clarity. Details on how box plots were generated are available at

https://matplotlib.org/stable/api/_as_gen/matplotlib.axes.Axes.boxplot.html#n

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Access to the PCAWG dataset can be obtained by applying for access at https://daco.icgc.org/. WGS data (both short- and long-read DNA sequencing) for the 786-O and Caki-1 cell lines have been deposited in NCBI with accession PRJNA868795.

Code availability

The code to filter loci from EHdn, termed TROPIC, is available on GitHub at https://github.com/AshwiniS7/TROPIC-Tandem-Repeat-Locus-Prioritization-in-Cancer. The code to perform local read depth normalization is available on GitHub at https://github.com/rashidalabri/cancer-rre-paper-analysis.

References

1. Hannan, A. J. Tandem repeats mediating genetic plasticity in health and disease. *Nat. Rev. Genet.* **19**, 286–298 (2018).

2. Gall-Duncan, T., Sato, N., Yuen, R. K. C. & Pearson, C. E. Advancing genomic technologies and clinical awareness accelerates discovery of disease-associated tandem repeat sequences. *Genome Res.* **32**, 1–27 (2022).

3. Hause, R. J., Pritchard, C. C., Shendure, J. & Salipante, S. J. Classification and characterization of microsatellite instability across 18 cancer types. *Nat. Med.* **22**, 1342–1350 (2016).

4. Cortes-Ciriano, I., Lee, S., Park, W. Y., Kim, T. M. & Park, P. J. A molecular portrait of microsatellite instability across multiple cancers. *Nat. Commun.* **8**, 15180 (2017).

5. Grünewald, T. G. P. et al. Chimeric EWSR1-FLI1 regulates the Ewing sarcoma susceptibility gene *EGR2* via a GGAA microsatellite. *Nat. Genet.* **47**, 1073–1078 (2015).

6. Aaltonen, L. A. et al. Clues to the pathogenesis of familial colorectal cancer. *Science* **260**, 812–816 (1993).

7. Thibodeau, S. N., Bren, G. & Schaid, D. Microsatellite instability in cancer of the proximal colon. *Science* **260**, 816–819 (1993).

8. Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. & Perucho, M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363**, 558–561 (1993).

9. Ho, S. S., Urban, A. E. & Mills, R. E. Structural variation in the sequencing era. *Nat. Rev. Genet.* **21**, 171–189 (2020).

10. Hannan, A. J. Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for ‘missing heritability’. *Trends Genet.* **26**, 59–65 (2010).

11. Wooster, R. et al. Instability of short tandem repeats (microsatellites) in human cancers. *Nat. Genet.* **6**, 152–156 (1994).

12. Risinger, J. I. et al. Genetic instability of microsatellites in endometrial carcinoma. *Cancer Res.* **53**, 5100–5103 (1993).

13. Han, H.-J., Yanagisawa, A., Kato, Y., Park, J.-G. & Nakamura, Y. Genetic instability in pancreatic cancer and poorly differentiated type of gastric cancer. *Cancer Res.* **53**, 5087–5089 (1993).

14. Panzer, S., Kuhl, D. P. A. & Caskey, C. T. Unstable triplet repeat sequences: a source of cancer mutations? *Stem Cells* **13**, 146–157 (1995).

15. Dolzhenko, E. et al. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res.* **27**, 1895–1903 (2017).

16. Dolzhenko, E. et al. ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. *Genome Biol.* **21**, 102 (2020).

17. Dashnow, H. et al. STRetch: detecting and discovering pathogenic short tandem repeat expansions. *Genome Biol.* **19**, 121 (2018).

18. Mousavi, N., Shleizer-Burko, S., Yanicky, R. & Gymrek, M. Profiling the genome-wide landscape of tandem repeat expansions. *Nucleic Acids Res.* **47**, e90 (2019).

19. Kristmundsdottir, S., Eggertsson, H. P., Arnadottir, G. A. & Halldorsson, B. V. popSTR2 enables clinical and population-scale genotyping of microsatellites. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btz913 (2019).

20. Tang, H. et al. Profiling of short-tandem-repeat disease alleles in 12,632 human whole genomes. *Am. J. Hum. Genet.* **101**, 700–715 (2017).

21. Rafehi, H. et al. Bioinformatics-based identification of expanded repeats: a non-reference intronic pentamer expansion in *RFC1* causes CANVAS. *Am. J. Hum. Genet.* **105**, 151–165 (2019).

22. Hannan, A. J. Repeat DNA expands our understanding of autism spectrum disorder. *Nature* **589**, 200–202 (2021).

23. Mitra, I. et al. Patterns of de novo tandem repeat mutations and their role in autism. *Nature* **589**, 246–250 (2021).

24. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).

25. Tankard, R. M. et al. Detecting expansions of tandem repeats in cohorts sequenced with short-read sequencing data. *Am. J. Hum. Genet.* **103**, 858–873 (2018).

26. Trost, B. et al. Genome-wide detection of tandem DNA repeats that are expanded in autism. *Nature* **589**, 80–86 (2020).

27. Tirkkonen, M. et al. Molecular cytogenetics of primary breast cancer by CGH. *Genes Chromosomes Cancer* **21**, 177–184 (1998).

28. Fujimoto, A. et al. Comprehensive analysis of indels in whole-genome microsatellite regions and microsatellite instability across 21 cancer types. *Genome Res.* **30**, 334–346 (2020).

29. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).

30. Sandi, C., Al-Mahdawi, S. & Pook, M. A. Epigenetics in Friedreich's ataxia: challenges and opportunities for therapy. *Genet. Res. Int.* **2013**, 852080 (2013).

31. Audano, P. A. et al. Characterizing the major structural variant alleles of the human genome. *Cell* **176**, 663–675 (2019).

32. Cavalcante, R. G. & Sartor, M. A. annotatr: genomic regions in context. *Bioinformatics* **33**, 2381–2383 (2017).

33. Moore, J. E. et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).

34. Pletscher-Frankild, S., Pallejà, A., Tsafou, K., Binder, J. X. & Jensen, L. J. DISEASES: text mining and data integration of disease–gene associations. *Methods* **74**, 83–89 (2015).

35. Schumacher, F. R. et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* **50**, 928–936 (2018).

36. Maor-Nof, M. et al. p53 is a central regulator driving neurodegeneration caused by *C9orf72* poly(PR). *Cell* **184**, 689–708 (2021).

37. Bae, B.-I. et al. p53 mediates cellular dysfunction and behavioral abnormalities in Huntington's disease. *Neuron* **47**, 29–41 (2005).

38. Sundararajan, R. & Freudenreich, C. H. Expanded CAG/CTG repeat DNA induces a checkpoint response that impacts cell proliferation in *Saccharomyces cerevisiae*. *PLoS Genet.* **7**, e1001339 (2011).

39. Lin, A., Zhang, J. & Luo, P. Crosstalk between the MSI status and tumor microenvironment in colorectal cancer. *Front. Immunol.* **11**, 2039 (2020).

40. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).

41. Barre, L. et al. Substrate specificity of the human UDP-glucuronosyltransferase UGT2B4 and UGT2B7. *FEBS J.* **274**, 1256–1264 (2007).

42. Rouleau, M. et al. Divergent expression and metabolic functions of human glucuronosyltransferases through alternative splicing. *Cell Rep.* **17**, 114–124 (2016).

43. Nakamori, M. et al. A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions in vivo. *Nat. Genet.* **52**, 146–159 (2020).

44. Erwin, G. S. et al. Synthetic transcription elongation factors license transcription across repressive chromatin. *Science* **358**, 1617–1622 (2017).

45. Erwin, G. S. G. S. et al. Synthetic genome readers target clustered binding sites across diverse chromatin states. *Proc. Natl Acad. Sci. USA* **113**, E7418–E7427 (2016).

46. Kim, J. C. & Mirkin, S. M. The balancing act of DNA repeat expansions. *Curr. Opin. Genet. Dev.* **23**, 280–288 (2013).

47. Russo, M. et al. Adaptive mutability of colorectal cancers in response to targeted therapies. *Science* **366**, 1473–1480 (2019).

48. Persi, E. et al. Proteomic and genomic signatures of repeat instability in cancer and adjacent normal tissues. *Proc. Natl Acad. Sci. USA* **116**, 16987–16996 (2019).

49. van Wietmarschen, N. et al. Repeat expansions confer WRN dependence in microsatellite-unstable cancers. *Nature* **586**, 292–298 (2020).

50. Edward, G. et al. The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. *Proc. Natl Acad. Sci. USA* **94**, 3320–3323 (1997).

51. Zhao, H., Nolley, R., Chan, A. M. W., Rankin, E. B. & Peehl, D. M. Cabozantinib inhibits tumor growth and metastasis of a patient-derived xenograft model of papillary renal cell carcinoma with *MET* mutation. *Cancer Biol. Ther.* **18**, 863–871 (2017).

52. McLean, C. Y. et al. GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

53. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

54. Seabold, S. & Perktold, J. Statsmodels: econometric and statistical modeling with Python. in *Proceedings of the 9th Python in Science Conference* (eds van der Walt, S. & Millman, J.) 92–96 (SciPy, 2010).

55. Kent, W. J. et al. The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

56. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

57. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).

58. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

59. Sun, H. et al. Quantitative integration of epigenomic variation and transcription factor binding using MAmotif toolkit identifies an important role of IRF2 as transcription activator at gene promoters. *Cell Discov.* **4**, 38 (2018).

60. Altshuler, D. M. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).

61. Pinese, M. et al. The Medical Genome Reference Bank contains whole genome and phenotype data of 2570 healthy elderly. *Nat. Commun.* **11**, 435 (2020).

62. Wu, S. et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* **575**, 699–703 (2019).

63. Fotsing, S. F. et al. The impact of short tandem repeat variation on gene expression. *Nat. Genet.* **51**, 1652–1659 (2019).

64. Dillon, L. W. et al. Production of extrachromosomal microDNAs is linked to mismatch repair pathways and transcriptional activity. *Cell Rep.* **11**, 1749–1759 (2015).

Acknowledgements

This work was supported by NIH grants U2CCA233311 (to M.P.S.) and K99HG011467 (to G.S.E.). G.S.E. was also supported by a Stanford Cancer Institute Postdoctoral Fellowship from the Ellie Guardino Research Fund. Computational support was provided by the Cancer Genomics Cloud (to G.G. and G.S.E.) and an AWS Cloud Research Grant (to G.S.E.). G.S.E. thanks P.S. Kim for early advice and encouragement. We thank C. Sabatti for advice on statistical analysis, S. O’Connor for preliminary help with data processing, K. Van Bortle for advice and L. Vanderploeg and M. Algama for figures. This work was also supported by a National Cancer Institute Cancer Center Support Grant (P30CA124435). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NCI.

Author information

Author notes

1. These authors contributed equally: Graham S. Erwin, Gamze Gürsoy

Authors and Affiliations

1. Department of Genetics, Stanford University, Stanford, CA, USA

   Graham S. Erwin, Rashid Al-Abri, Ashwini Suriyaprakash, Kevin Zhu, Shannon M. White, Lucia Ramirez, Ananya Vadlakonda, Alekhya Vadlakonda, Konor von Kraut, Julia Park, Charlotte M. Brannon, Daniel A. Sumano, Raushun A. Kirtikar & Michael P. Snyder

2. Department of Biomedical Informatics, Columbia University, New York, NY, USA

   Gamze Gürsoy

3. New York Genome Center, New York, NY, USA

   Gamze Gürsoy

4. Illumina, Inc., San Diego, CA, USA

   Egor Dolzhenko & Michael A. Eberle

5. Division of Oncology, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA

   Christian R. Hoerner, Thomas J. Metzner & Alice C. Fan

6. Data Science Program, Northwestern University, Chicago, IL, USA

   Alicia A. Erwin

7. Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada

Ryan K. C. Yuen

8. Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

Ryan K. C. Yuen

9. Department of Urology, Stanford University School of Medicine, Stanford, CA, USA

Alice C. Fan & John T. Leppert

10. Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, USA

John T. Leppert

11. Division of Nephrology, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA

John T. Leppert

12. Computational Biology and Bioinformatics Program, Yale University, New Haven, CT, USA

Mark Gerstein

13. Molecular Biophysics and Biochemistry Department, Yale University, New Haven, CT, USA

Mark Gerstein

14. Department of Computer Science, Yale University, New Haven, CT, USA

Mark Gerstein

Contributions

G.S.E. conceived the study. G.S.E., G.G., A.C.F., J.T.L., M.A.E., M.P.S. and M.G. supervised research. G.S.E., G.G., R.A.-A., A.S., E.D., J.P., C.M.B., K.Z., R.K.C.Y. and A.A.E. analysed data. G.S.E., C.R.H., L.R., Ananya Vadlakonda, Alekhya Vadlakonda, K.v.K., R.A.K., D.A.S., S.M.W. and T.J.M. conducted wet lab experiments. G.S.E. and M.P.S. wrote the manuscript with input from all the authors.

Corresponding authors

Correspondence to Graham S. Erwin, Mark Gerstein or Michael P. Snyder.

Ethics declarations

Competing interests

G.S.E. and M.P.S. are inventors on a patent application describing anti-proliferative agents. E.D. and M.A.E. are shareholders and are currently or were formerly employed by Illumina and Pacific Biosciences. The other authors declare no competing interests.

Peer review

Peer review information

Nature thanks Anthony Hannan, Matthew Disney and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Overview of PCAWG data and analysis with ExpansionHunter De Novo.

a, Distribution of cancer genomes analysed across 29 human cancers in the PCAWG data. **b**, Distribution of p-values following candidate recurrent repeat expansion (rRE) analysis with ExpansionHunter Denovo (one-sided Wilcoxon rank-sum test).

Extended Data Fig. 2 Benchmarking EHdn.

a, Comparison of anchored in-repeat reads (IRRs) to long-read sequencing reads. Long-read sequencing confirmation rate across all tandem repeats (TRs, motifs 2–20 bp), short TRs with motifs from 2–6 bp, and variable number TRs with motifs from 7–20 bp. **b**, Confirmation rate versus number of anchored IRRs. **c**, Effect of downsampling on the identification of the rRE in the intron of *UGT2B7* in kidney cancer. Tumour genomes from the PCAWG dataset were downsampled to the specified number. ExpansionHunter De Novo was run, and the resulting Bonferroni-correct p-value is depicted for the given sequencing coverage. Corrected p-value from one-sided Wilcoxon rank-sum test with Bonferroni correction. **d**, Estimation of the frequency of repeat expansions in rRE loci in the general population. The number of rREs (count) corresponding to each bin is plotted on the y-axis. Results are from analysis of 1000 Genomes Project samples[60] (n = 2,504) (GRCh38) and Medical Genome Reference Bank[61] samples (n = 4,010).

Extended Data Fig. 3 Local read depth normalization of recurrent repeat expansion (rRE) candidates.

a, Examples of read depth before and after local normalization. **b**, Examples of anchored in-repeat read (IRRs) before and after local normalization. The read depth for the locus on the left is derived from TCGA data, and the read depth for the locus on the right is derived from PCAWG data. Q-values were calculated from two-tailed Student's t-test with FDR correction by Benjamini-Hochberg. FDR q-value=4.83e-05 and 0.54 for Kidney-RCC and

Breast-AdenoCA, respectively (n = 74 Kidney-RCC genomes and n = 193 Breast-AdenoCA genomes analyzed). **c**, Workflow to identify rREs. **d**, Detection rate in an independent cohort of samples.

Extended Data Fig. 4 Benchmarking LRDN and EHdn.

a,b, Benchmarking the local read depth normalization filter (n = 10 loci analysed). **c**, The anchored IRR quotient was calculated as (tumour anchored IRR – normal anchored IRR)/(normal anchored IRR + 1). Dashed line at 2.5 indicates the threshold for calling a locus as a repeat expansion in a cancer genome. **d**, ExpansionHunter was used to estimate repeat sizes from short-read sequencing data, and the results were visualized with REViewer (see Methods). The allele with the longest repeat tract for normal and tumour samples is shown. The TR is depicted in red, and the flanking regions are depicted in blue.

Extended Data Fig. 5 Association of rREs with genetic features.

a, Correlation of rREs with MSI-High cancers. **b**, **c**, Association of rREs with mutational signatures. **b**, Correlation between DBS2 and the number of rREs detected. **c**, Correlation between DBS2 and the number of rREs detected when Lung-SCC data are omitted from the analysis.

Extended Data Fig. 6 Distribution of rREs across the genome.

a, Distance of rREs to the nearest centromere or telomere. **b**, Distribution of rREs across early- and late-replicating regions of the genome. Welch's *t*-test (two-tailed, not significant). **c**, Circos plot depicting (from outside to inside) p-value of rREs, location of rREs where darker shading indicates the rRE observed across 3 cancers, early and late replicating regions (yellow and purple, respectively), and simple sequence repeats. This plot depicts the overlay between different data types and the distribution of rREs across the genome.

Extended Data Fig. 7 Molecular features of rREs.

a, Overlap of rREs with other datasets. The fraction of rREs overlapping with other catalogues of TRs and genomic instability. From left to right in the figure, recurrently altered STRs in cancer (Supplementary Data 14 from ref. [4]; PMID: 28585546), extrachromosomal circular DNA (ecDNA, circular amplification events from Supplementary Table 1; ref. [62]; PMID: 31748743), unstable STRs in cancer (Supplementary Table 10 from ref. [3]; PMID: 27694933), eSTRs (Supplementary Data 1; ref. [63]; PMID: 31676866), and microDNA (From C4-2, ES2, LNCaP, OVCAR8, and PC-3 cells; ref. [64]; PMID: 26051933). The PubMed ID for each corresponding manuscript is included in the figure. For the overlap of rREs with microDNA, we looked at loci that we attempted to detect in an independent cohort of cancer samples, and we found that we tested 11 loci. Of the 11 rREs tested, 8 (72%) were detected in the independent cohort of cancer samples. **b**, Distribution of rRE motif length across cancer types. **b**,**c**, Association of rREs with regulatory elements. **b**, Distance of simple sequence repeats and rREs to the nearest candidate cis-regulatory elements (cCREs). Key: promoter-like signature (P), proximal enhancer-like signature (p), distal enhancer-like signature (d), DNase-H3K4me3 (D), and CTCF-only (C). **c**, Signal tracks depicting rREs near regulatory elements (n = 950,091 simple repeats and n = 160 rREs). **d**, Association between rREs in prostate cancer and risk loci in prostate cancer. Signal trace showing an rRE detected in prostate cancer and a risk locus for prostate cancer.

Extended Data Fig. 8 Analysis of cytotoxic activity.

a, Analysis of *UGT2B7* GAAA rRE in patients with clear cell RCC. N, normal tissue; T, tumour tissue. For gel source data, see Fig. S1. **b**, *UGT2B7* in RCC patients. **b**, Expression of *UGT2B7* (transcripts per million, TPM) in RCC samples as a function of the detection of the rRE in *UGT2B7*. P value computed with Welch's *t*-test (two-tailed. **c**, Kaplan-Meier survival plots of RCC patients stratified by rRE in the intron of *UGT2B7*. P value computed with Welch's *t*-test (two-tailed).

Extended Data Fig. 9 Association of rREs with cytotoxic activity.

P values computed with Welch's *t*-test (two-tailed) with FDR correction (Benjamini-Hochberg) ($n = 49$ Kidney-RCC genomes and $n = 85$ Ovary-AdenoCA genomes analysed).

Extended Data Fig. 10 Syn-TEF treatment of RCC cell lines.

a, Quantitation of the percentage of propidium iodide-positive cells. P values are from a one-way ANOVA adjusted with Bonferroni correction for multiple comparisons. Results are mean ± s.e.m. ($n = 3$ biological replicates, except $n = 2$ biological replicates for Syn-TEF3 in 786-O). **b**, Live cell microscopy of Caki-1 and 786-O cells stained with propidium iodide (red) and Hoechst 33342 (blue). Scale bars, 100 µm. **c**, Relative cell density of RCC cell lines following treatment (72 h) with compounds (50 µM Syn-TEF or 0.1% DMSO vehicle, as indicated). Results are mean ± s.e.m. (ACHN and RCC-4 are $n = 4$ biological replicates, A498 and Caki-2 are $n = 3$ biological replicates).

Supplementary information

Supplementary Information

This file contains Supplementary Figs. 1 and 2.

Reporting Summary

Supplementary Tables

This file contains Supplementary Tables 1–7.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Erwin, G.S., Gürsoy, G., Al-Abri, R. *et al.* Recurrent repeat expansions in human cancer genomes. *Nature* **613**, 96–102 (2023). https://doi.org/10.1038/s41586-022-05515-1

- Received: 13 August 2021
- Accepted: 02 November 2022
- Published: 14 December 2022
- Issue Date: 05 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05515-1

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

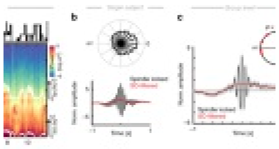
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

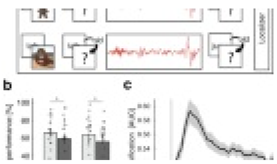
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05515-1



Bidirectional prefrontal-hippocampal dynamics organize information transfer during sleep in humans

08 August 2019

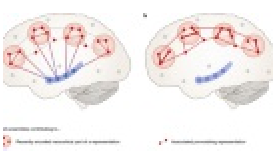
Randolph F. Helfrich, Janna D. Lendner, … Robert T. Knight



Endogenous memory reactivation during sleep in humans is clocked by slow oscillation-spindle complexes

25 May 2021

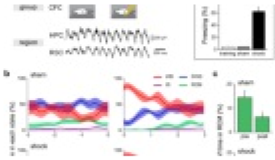
Thomas Schreiner, Marit Petzka, … Bernhard P. Staresina



Mechanisms of systems memory consolidation during sleep

26 August 2019

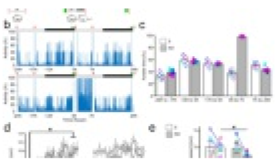
Jens G. Klinzing, Niels Niethard & Jan Born



Hippocampus-retrosplenial cortex interaction is increased during phasic REM and contributes to memory consolidation

22 June 2021

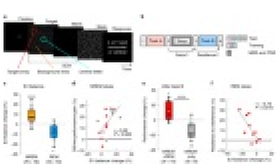
Daniel Gomes de Almeida-Filho, Bruna Del Vechio Koike, … Claudio Marcos Queiroz



Net decrease in spine-surface GluA1-containing AMPA receptors after post-learning sleep in the adult mouse cortex

17 May 2021

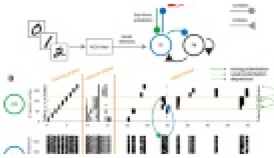
Daisuke Miyamoto, William Marshall, … Chiara Cirelli



Complementary contributions of non-REM and REM sleep to visual learning

20 July 2020

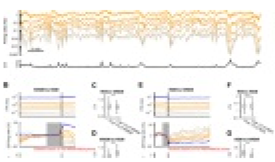
Masako Tamaki, Zhiyan Wang, … Yuka Sasaki



[Sleep-like slow oscillations improve visual classification through synaptic homeostasis and memory association in a thalamo-cortical model]

20 June 2019

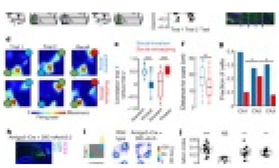
Cristiano Capone, Elena Pastorelli, … Pier Stanislao Paolucci



[Neuronal firing rates diverge during REM and homogenize during non-REM]

24 January 2019

Hiroyuki Miyawaki, Brendon O. Watson & Kamran Diba



[Hippocampal CA2 sharp-wave ripples reactivate and promote social memory]

23 September 2020

Azahara Oliva, Antonio Fernández-Ruiz, … Steven A. Siegelbaum

- Article
- Open Access
- Published: 14 December 2022

Cortical–hippocampal coupling during manifold exploration in motor cortex

- Jaekyung Kim,
- Abhilasha Joshi,
- Loren Frank &
- …
- Karunesh Ganguly

Show authors

Nature **volume 613**, pages 103–110 (2023)

- 11k Accesses
- 259 Altmetric
- Metrics details

Abstract

Systems consolidation—a process for long-term memory stabilization—has been hypothesized to occur in two stages[1,2,3,4]. Whereas new memories require the hippocampus[5,6,7,8,9], they become integrated into cortical networks over time[10,11,12], making them independent of the hippocampus. How hippocampal–cortical dialogue precisely evolves during this and how

cortical representations change in concert is unknown. Here, we use a skill learning task[13,14] to monitor the dynamics of cross-area coupling during non-rapid eye movement sleep along with changes in primary motor cortex (M1) representational stability. Our results indicate that precise cross-area coupling between hippocampus, prefrontal cortex and M1 can demarcate two distinct stages of processing. We specifically find that each animal demonstrates a sharp increase in prefrontal cortex and M1 sleep slow oscillation coupling with stabilization of performance. This sharp increase then predicts a drop in hippocampal sharp-wave ripple (SWR)–M1 slow oscillation coupling—suggesting feedback to inform hippocampal disengagement and transition to a second stage. Notably, the first stage shows significant increases in hippocampal SWR–M1 slow oscillation coupling in the post-training sleep and is closely associated with rapid learning and variability of the M1 low-dimensional manifold. Strikingly, even after consolidation, inducing new manifold exploration by changing task parameters re-engages hippocampal–M1 coupling. We thus find evidence for dynamic hippocampal–cortical dialogue associated with manifold exploration during learning and adaptation.

Main

Memory stabilization through systems consolidation has long been hypothesized to follow a two-stage process[1,2,3,4]. Although both classic and recent studies provide support for a hippocampus-dependent and an independent phase[3,6,15,16], it is unclear how cortical representations evolve with hippocampal–cortical dialogue. Moreover, it is unknown at what timescales such coordination occurs and what processes might inform transitions between stages during systems consolidation.

We use a prehension skill learning task to monitor hippocampal–cortical (prefrontal cortex (PFC) and primary motor cortex (M1)) dialogue across both sleep and awake states. Because prehension requires M1 (refs. [14,17]), we can monitor consolidation in an essential region. Reactivation of task ensembles in M1 during non-rapid eye movement sleep (NREMS) is also known to be essential for skill learning[18,19,20]. Furthermore, global sleep slow oscillations (SOs), a phenomenon linked to the PFC and

hippocampus[1,21,22,23], appear to be critical drivers of local reactivation in M1 (ref. [18]). Together, this suggests that consolidation of motor memories may involve dialogue between cortex and hippocampus during NREMS.

Although classic studies indicated that ‘procedural memories’ are hippocampus-independent[24], recent functional MRI studies have found evidence for hippocampal activation in very early phases[25,26]. The nature of this hippocampal activation, however, remains unclear, and specifically whether hippocampal sharp-wave ripples (SWRs)—a key neurophysiological marker of memory consolidation during NREMS[1,27,28]—are associated with skill learning remains unknown. We therefore tested the hypothesis that time-varying changes in the coupling between SOs and SWRs during NREMS demarcate stages of consolidation.

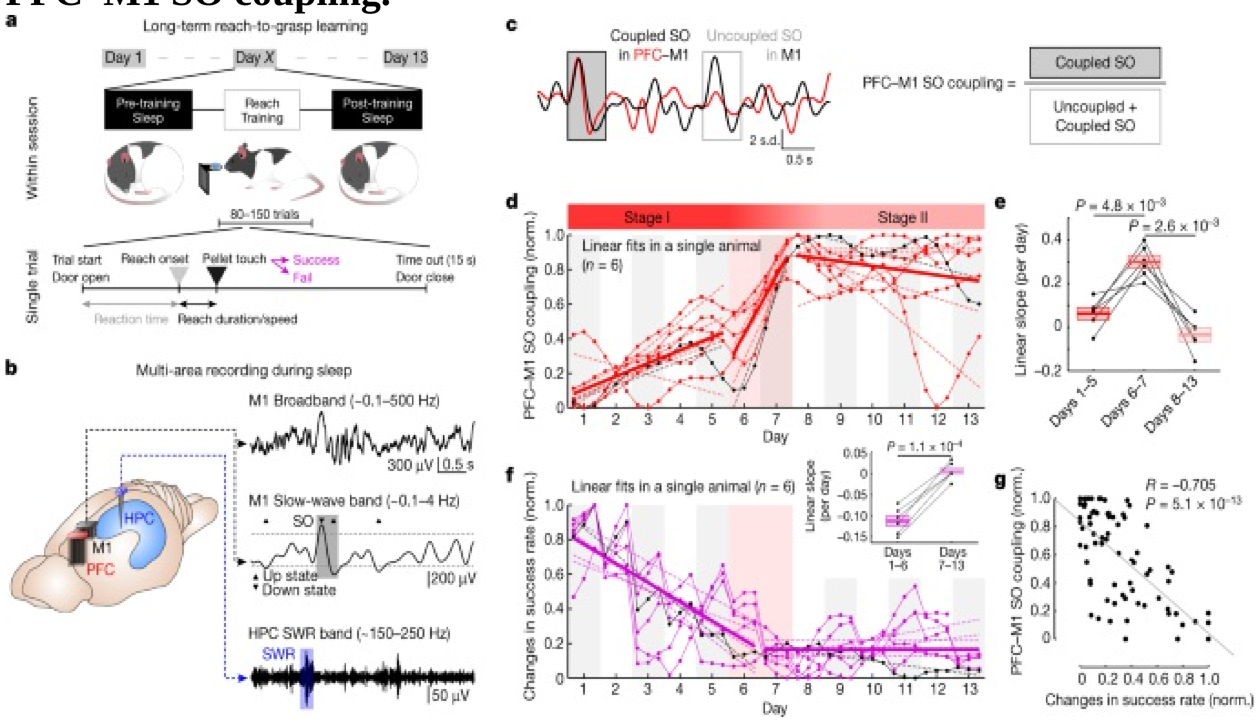
A closely related hypothesis is that each of these two stages is associated with distinct phases of consolidation in M1. The temporal evolution of neural population dynamics (‘neural trajectories’) within a low-dimensional manifold (patterns of shared variance) is closely tied to skilled performance[29,30,31]. Whereas neural trajectories are stable for well-practised behaviours, they are variable during learning[32,33]. We quantified neural trajectory variance using a measure of fidelity. We expect low fidelity (large variance) during early learning because of ‘manifold exploration’. By contrast, during ‘manifold consolidation’, we expect increasing fidelity. It is unclear how systems consolidation might relate to manifold exploration versus consolidation.

Here, we demonstrate that changes in PFC–M1 SO coupling are closely linked to changes in performance across days and can both predict a drop in SWRs’ learning-related changes in M1 and demarcate a transition away from manifold exploration. Strikingly, there is a sharp increase in PFC–M1 SO coupling that coincides with stabilization of performance and manifold consolidation. This sharp increase in coupling is also a strong temporal predictor of the disengagement of increases in M1 SWR coupling; manifold consolidation continued to occur after this phase. Interestingly, inducing errors re-engaged manifold exploration and hippocampal–cortical coupling, further supporting that hippocampal–cortical dialogue may support manifold exploration.

PFC–M1 SO coupling over long-term skill learning

If cortical interactions during NREMS play a role in consolidation, then there should be changes in the coordination of SOs in PFC and M1 with learning. We measured performance in a reach-to-grasp task over approximately 13 d ($n = 6$ rats; Fig. 1a), while also monitoring SOs in M1 and PFC as well as hippocampal SWRs (Fig. 1b and Extended Data Fig. 1). We first examined changes in the precise temporal coupling between SOs in PFC and M1. 'PFC–M1 SO coupling' is the number of M1 SOs occurring within 200 ms from a PFC SO divided by the total number of SOs in M1 (Fig. 1c). We examined changes in PFC–M1 SO coupling within and across sessions, particularly as there may be slow changes over days[33,34]. Indeed, PFC–M1 SO coupling demonstrated slow changes across days, but values in 'pre-training' and the 'post-training' sleep were similar within a day (Extended Data Fig. 2a). We thus focused only on the post-training sleep to examine PFC–M1 SO coupling over long-term learning. Notably, we found significantly higher PFC–M1 SO coupling in the late compared with the early period (Fig. 1d; normalization based on the minimum and maximum values across all days; 'early' is days 1–4 and 'late' is days 10–13).

Fig. 1: Changes in performance inversely correlated with increase in PFC–M1 SO coupling.



a, Flow chart of reach-to-grasp task training experiment. **b**, Examples of the broadband (0.1–500 Hz) and the filtered local-field potential (LFP) trace in M1 for SOs (0.1–4 Hz) and in hippocampus (HPC) for SWRs (150–250 Hz) during sleep. SOs and SWRs are marked by grey and blue boxes, respectively. Horizontal dashed lines indicate the threshold detecting SOs up-/down-states and SWRs onset. **c**, Schematic showing measurement of temporal coupling of M1 SOs from PFC SOs, that is, PFC–M1 SO coupling. **d**, Time courses of PFC–M1 SO coupling. Lines represent piecewise linear regression fits. Piecewise linear regression fits are shown with dashed lines for the fits in a single animal ($n = 6$ animals) and with the solid line using all six animals. In each animal, scale of minimum-to-maximum was normalized to range from 0 to 1; same as **f** (details in Methods). **e**, Comparison of linear slopes across three periods ($n = 6$ animals); one-way analysis of variance (ANOVA), $F_{12,15} = 33.39$, $P = 3.0 \times 10^{-6}$; post hoc two-sided paired *t*-test, corrected for multiple comparison, days 1–5 versus days 6–7: $P = 4.8 \times 10^{-3}$, days 6–7 versus days 8–13: $P = 2.6 \times 10^{-3}$. Mean ± s.e.m. **f**, Time courses of changes in success rate (based on 2 d history; Extended Data Fig. 2e). Piecewise linear regression fits are shown with dashed lines for the fits in a single animal ($n = 6$ animals) and with the solid line using all six animals. Inset: linear slopes comparison ($n = 6$ animals); two-sided paired *t*-test, $t_5 = -10.93$, $P = 1.1 \times 10^{-4}$. Mean ± s.e.m. **g**, Relationship between changes in success rate and PFC–M1 SO coupling. Across six rats, PFC–M1 SO coupling was well predicted by changes in success rate using a linear regression fit. Norm., normalized.

We next examined the temporal dynamics of the PFC–M1 SO coupling. We used moving windows of varying lengths (3–7 d) to identify rates of change and the location of rapid transitions (Methods and Extended Data Fig. 2c). We found a sharp transition around day 6 in all animals. On this basis, we measured linear slopes for the three periods and found a significantly fast transition on days 6–7 (Fig. 1e). After this period, there was elevated PFC–M1 SO coupling. There was no significant change in the rates of M1 SOs (Extended Data Fig. 1e). Although there was a significant increase in the rates of PFC SOs, this could not account for changes in coupling. First, when we generally shuffled the timing of PFC and M1 SOs, we did not observe this change (Extended Data Fig. 2b). Second, when we ‘subsampled’ the number of PFC SOs to match rates across sessions, we still observed nearly identical

increases in coupling (Extended Data Fig. 2d). Finally, total recorded NREMS periods were similar in both early and late (days 1–4, $n = 24$ sessions in 6 rats: 53.35 ± 2.34 min versus days 10–13, $n = 24$ sessions in 6 rats: 49.03 ± 2.29 min, linear mixed-effects model (LME) with two-sided t-test, $t_{46} = -1.61$, $P = 0.11$). Thus, there was more precise coupling of PFC–M1 SOs with learning.

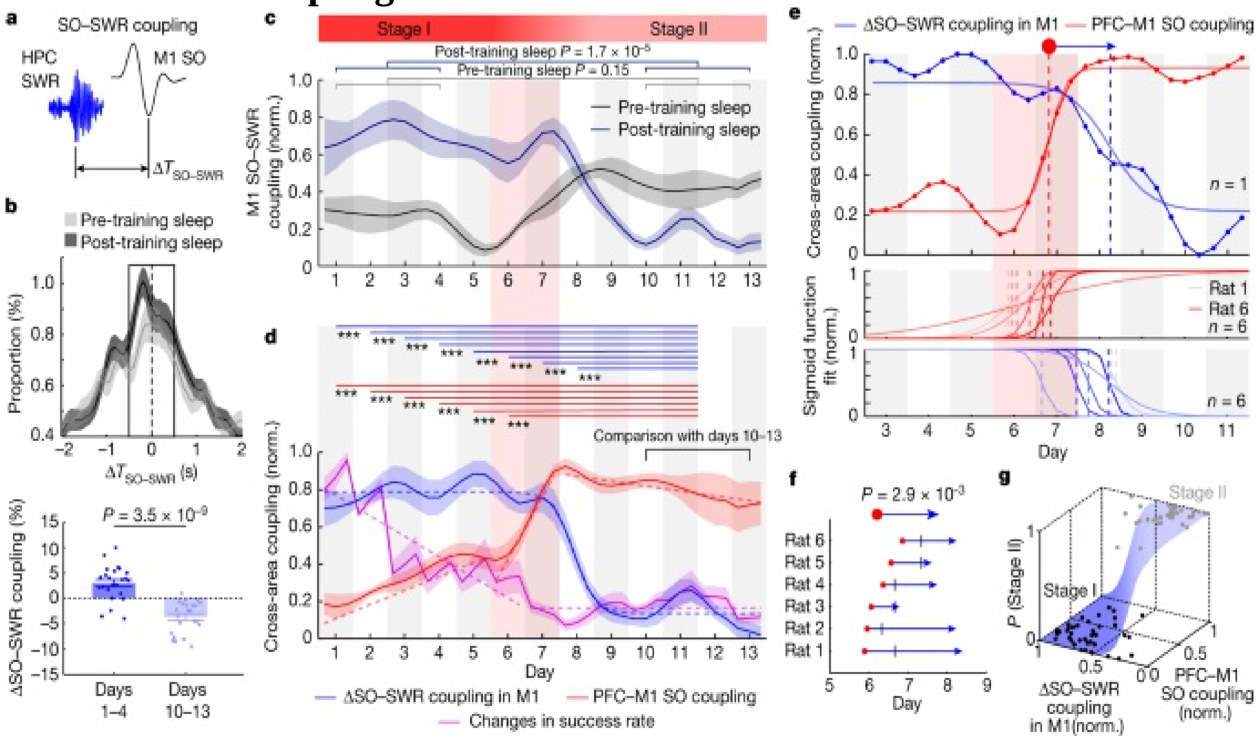
The transition from low to high PFC–M1 SO coupling occurred as the success rate stabilized (Fig. 1f,g; example shown in Extended Data Fig. 2e,f). To examine the rates of performance change before and after the transition day, we fit a line to the changes in success rate for each animal for days 1–6 and days 7–13. This revealed an initial reduction in the change of success rate (behavioural exploration); this was followed by stable performance. This indicates that the transition from exploration to stable performance was temporally linked to a sharp increase in PFC–M1 SO coupling.

Stepwise drop in M1 SO–SWR coupling

We next wondered whether the increased PFC–M1 coordination is temporally associated with changes in hippocampus–M1 coordination, as would be predicted for a memory consolidation process in M1 that evolves to become hippocampus-independent[1,2,3,4,35]. We first quantified coordination between hippocampal SWRs (Fig. 1b) and M1 SOs by measuring the time lag between an M1 SO and its nearest SWR ($\Delta T_{\text{SO–SWR}}$; Fig. 2a). We measured ‘SO–SWR coupling’ (probability of SWRs −0.75 to +0.75 s from an SO; black box in Fig. 2b and Methods) in the pre- and post-training sleep. In comparison with both a pre-training sleep and a shuffled distribution, the post-training sleep in early learning (days 1–4) had a significantly greater incidence of SWRs with precise temporal proximity to M1 SOs (Fig. 2b and Extended Data Fig. 3a). Notably, whereas PFC–M1 SO coupling demonstrated a slow change over days, SO–SWR coupling showed a faster learning-related change, that is, strong changes from pre- to post-training sleep (Fig. 2c). These results are perhaps consistent with views of the hippocampus and the cortex being, respectively, fast and slow learning systems[2].

Fig. 2: Sharp increase in PFC–M1 SO coupling predicts stepwise drop in

SWR–M1 SO coupling.



a, SO–SWR coupling in M1: temporal lags of SWRs from M1 SO, $\Delta T_{\mathrm{SO–SWR}}$. **b**, Comparison of M1 SO–SWR coupling (box: probability of SWR in −0.75 to +0.75 s window relative to SO). Top, $\Delta T_{\mathrm{SO–SWR}}$ distribution during pre- ($n = 24$ sessions in 6 rats, also hereafter) and post-training sleep of early. SO–SWR coupling in window was higher in post-training sleep (two-sided paired *t*-test, $t_{23} = -2.84$, $P = 9.3 \times 10^{-3}$). Bottom, ΔSO–SWR coupling change of SO–SWR (pre- versus post sleep) higher during early (days 1–4) versus late (days 10–13); LME, two-sided *t*-test, $t_{46} = -7.28$, $P = 3.5 \times 10^{-9}$. Mean ± s.e.m., also hereafter. **c**, Average time course of M1 SO–SWR coupling over the motor learning during pre-training sleep (early: 22.71 ± 0.98 versus late: 24.35 ± 1.05, LME, two-sided *t*-test, $t_{46} = 1.61$, $P = 0.15$) and post-training sleep (early: 24.95 ± 0.85 versus late: 20.47 ± 0.83, LME, two-sided *t*-test, $t_{46} = -4.81$, $P = 1.7 \times 10^{-5}$). The time course in pre-training sleep was normalized, referencing to the post-training sleep in each animal. **d**, Average time course of PFC–M1 SO coupling (red), ΔSO–SWR coupling in M1 (blue) and changes in success rate (magenta). Thick dashed sigmoid curve indicates sigmoid function fit. Thick dashed linear lines indicate piecewise linear regression fits reproduced from Fig. 1d,f. Top

horizontal lines: significant difference from the days 10–13; PFC–M1 SO coupling: one-way ANOVA, $F_{12,218} = 44.66$, $P = 5.0 \times 10^{-52}$; post hoc LME, two-sided *t*-test, corrected for multiple comparison, ***$P < 5.2 \times 10^{-5}$; ΔSO–SWR coupling: one-way ANOVA, $F_{12,218} = 29.30$, $P = 4.6 \times 10^{-39}$; post hoc LME, two-sided *t*-test, corrected for multiple comparison, ***$P < 2.2 \times 10^{-6}$. **e**,**f**, Comparison of stepwise increase of PFC–M1 SO coupling with drop in ΔSO–SWR coupling in single animal (**e**, top); middle and bottom, sigmoid function fits for each animal. Sharpest changes marked by red circles and blue arrowheads. Significant for six animals; two-sided paired *t*-test, $P = 2.9 \times 10^{-3}$ (**f**); black bars represent stage transitions using change-point analysis. **g**, Probability of stage II predicted using logistic regression as function of ΔSO–SWR coupling in M1 and PFC–M1 SO coupling (blue surface). 'Stage I' (black) and 'Stage II' (grey) represent true sessions before and after grand-midpoint (median of red dot and blue arrowhead in an individual animal in **f**).

To quantify changes in hippocampal–M1 coordination, we measured 'ΔSO–SWR coupling' in M1 as the normalized difference of coupling between post- and pre-training sleep. There was significantly stronger ΔSO–SWR coupling during the early versus the late period (Fig. 2b and Extended Data Fig. 3b). The temporal dynamics of ΔSO–SWR coupling in M1 also showed stepwise dynamics (Fig. 2d, sigmoid function fit, $R^2 = 0.97$, $P = 1.6 \times 10^{-46}$). There was a marked drop slightly after the sharp increase in PFC–M1 SO coupling (days 6–7). This was not the case if we disrupted the SWR temporal structure by shuffling across time (Extended Data Fig. 3c). This was also not due to a difference in SWR rates (Extended Data Fig. 1e). Notably, the ΔSO–SWR coupling in PFC showed sustained higher values and a delayed drop (Extended Data Fig. 3d).

To examine temporal links between cross-area coupling transitions, we fit a sigmoid function to each animal's cross-area coupling (thick curves in Fig. 2e and Methods). Across six rats, the stepwise dynamics were well predicted by a sigmoid function in ΔSO–SWR coupling in M1 (fit per animal, $R^2 = 0.66 \pm 0.078$, $P < 3.27 \times 10^{-4}$) and in PFC–M1 SO coupling (fit per animal, $R^2 = 0.86 \pm 0.056$, $P < 4.8 \times 10^{-19}$). This increase in PFC–M1 SO coupling—as measured by the midpoint of the sigmoid function fit, that is, the red dot—was followed by a sharp drop in ΔSO–SWR coupling in M1 (blue

arrowhead). This lag was significant across the six animals (Fig. 2f; rise time in PFC–M1 SO coupling, that is, red dots, $n = 6$ rats: 6.2 ± 0.16 d versus drop time in ΔSO–SWR coupling in M1, that is, blue arrowheads, $n = 6$ rats: 7.9 ± 0.23 d; two-sided paired t-test, $t_5 = -5.4$, $P = 2.9 \times 10^{-3}$). We arrived at a very similar conclusion when estimating the transition using 'change-point analysis' (black bars in Fig. 2f, Extended Data Fig. 3e and Methods).

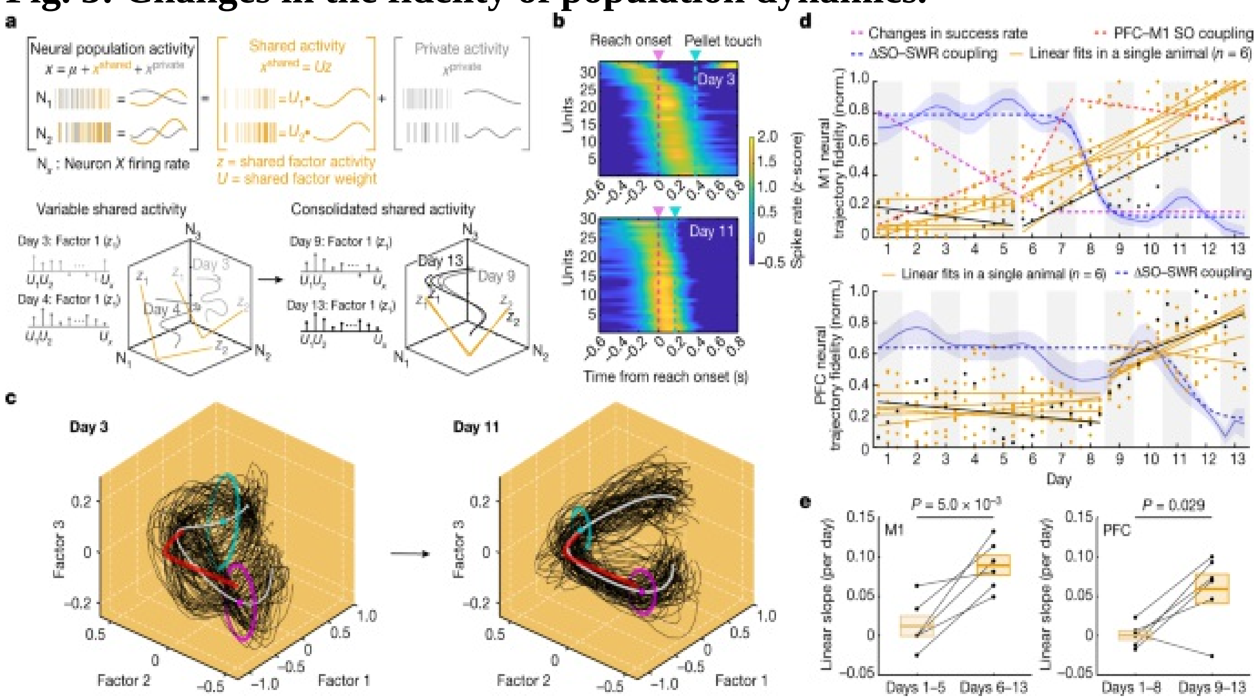
These findings indicate that the full engagement of NREMS PFC–M1 coordination predicts hippocampal–M1 disengagement. The median of the rise time in PFC–M1 coordination and the drop time in hippocampal–M1 coordination (hereafter, grand-midpoint) suggest two distinct stages of consolidation, hereafter termed stage I or stage II. We thus assessed whether knowledge about hippocampus, PFC and M1 states could identify whether a given animal was in stage I or stage II. We quantified this using a logistic regression model of ΔSO–SWR coupling in M1 and PFC–M1 SO coupling to predict whether a given session was stage I or II. The probability of stage II was significant using this model ($P = 4.3 \times 10^{-3}$ for ΔSO–SWR coupling in M1 and $P = 2.4 \times 10^{-3}$ for PFC–M1 SO coupling; Fig. 2g). Thus, our results indicate that changes in PFC–M1 coordination and hippocampal–M1 disengagement can define two distinct phases during systems consolidation.

Memory consolidation is known to require the precise coupling of sleep spindles to SOs and SWRs[18,36,37]. We examined the temporal relationship of spindles to SOs and SWRs. Our analyses suggest that whereas cross-area communication is related to SOs (possible orchestration by SWRs), spindle activity appears to be a local process (model illustrated in Extended Data Fig. 4a). We did not observe changes in 'triple-coupling' (that is, precise coupling of SWR, SO and spindle; Extended Data Fig. 4b). However, we found learning-related changes in the local coupling of spindles with M1 SOs and PFC SOs separately (Extended Data Fig. 5). There were stronger task-dependent changes in the SO–Spindle coupling during the early versus the late period (that is, pre- versus post-training sleep, ΔSO–Spindle coupling[18,38]). This suggest a model in which changes in area-specific SO–Spindle coupling are important. However, spindles do not appear to demonstrate a specific change over days during coordinated SWR and SO events.

Manifold consolidation with drop in SO–SWR coupling

We investigated how these cross-area interactions relate to changes in awake task-related activity during learning. On the basis of a model that early skill learning is associated with manifold exploration and later learning is associated with the stabilization of population dynamics (manifold consolidation)[29,32,39,40], we monitored the stability of population dynamics in M1 and PFC using Gaussian process factor analysis (GPFA). GPFA identifies a low-dimensional cortical manifold which represents patterns of neural co-firing (that is, shared activity; Fig. 3a)[29,32,33]. We calculated GFPA neural trajectories for each day and focused on the top three factors during reaching (Fig. 3b; see Methods for alignment across days). First, we focused on M1. As expected, neural trajectories were variable across trials in a representative early session, whereas they were more stereotyped in a late session (Fig. 3c and Extended Data Fig. 6a–d). Consistency across days, particularly for periods of stable performance, is also consistent with recent findings of stable manifolds across days[32,41].

Fig. 3: Changes in the fidelity of population dynamics.



a, Schematic diagram showing shared and private activity of neural population activity using GPFA. **b**, Examples of PETH in M1 during reach

period at days 3 and 11; locked to the reach onsets. Durations from reach onset to pellet touch were typically longer at the early compared with the late period of learning. **c**, Examples of neural trajectories (black) travelling on the shared activity space defined by the top three GPFA factors in M1. Grey curves, mean template. Red curves, 'optimal' neural trajectory for a successful task. Dots, mean reach onset (magenta) and pellet touch (cyan) times. Ellipses, 90% confidence area of covariance. **d**, Neural trajectory fidelity over learning in M1 (top) and PFC (bottom); each dot represents measurement in a tertile of single-session trials. Lines represent linear regression fits in each animal for either days 1–5 or days 6–13 in M1 and days 1–8 or days 9–13 in PFC. Top: red and magenta dashed lines represent the piecewise linear regression fits of the changes in PFC–M1 SO coupling from Fig. 1d and in success rate from Fig. 1f, respectively. Blue curves represent ΔSO–SWR coupling from Fig. 2d and Extended Data Fig. 3d. **e**, Comparison of the slopes in **d** between two periods (two-sided paired *t*-test, $n = 6$ rats, $P = 5.0 \times 10^{-3}$ for M1, $P = 0.029$ for PFC). Transition period was determined by varying windows (Extended Data Fig. 6e). Mean ± s.e.m.

We next quantified the changes in trajectory stability over days. To do so, we calculated a 'GPFA correlation (*R*)', defined as the correlation coefficient of an individual neural trajectory with the average neural trajectory of the final 3 d associated with the best performance (Methods). We then calculated a neural trajectory 'fidelity', a signal-to-noise ratio[42,43,44] of GPFA *R* values across days; it captured the consistency of GPFA *R* values across trials. The dots (orange for $n = 5$ and black for a single animal) in Fig. 3d show trends towards increased fidelity.

On the basis of the observed changes in PFC–M1 SO coupling (Fig. 1d), we wondered whether there might be two distinct stages for activity patterns within the M1 manifold. We again used moving windows (for example, 3–7 d) to estimate slopes to objectively determine the greatest change in M1 fidelity; we found that between days 5 and 6 was a consistent transition point (Extended Data Fig. 6e). This suggested that days 1–5 and days 6–13 represented two periods with distinct rates of change. We thus fit a line for each animal for days 1–5 and then days 6–13. This demarcation revealed an essentially flat slope of approximately zero for days 1–5 (manifold exploration) and then a gradual linear increase trajectory fidelity (manifold

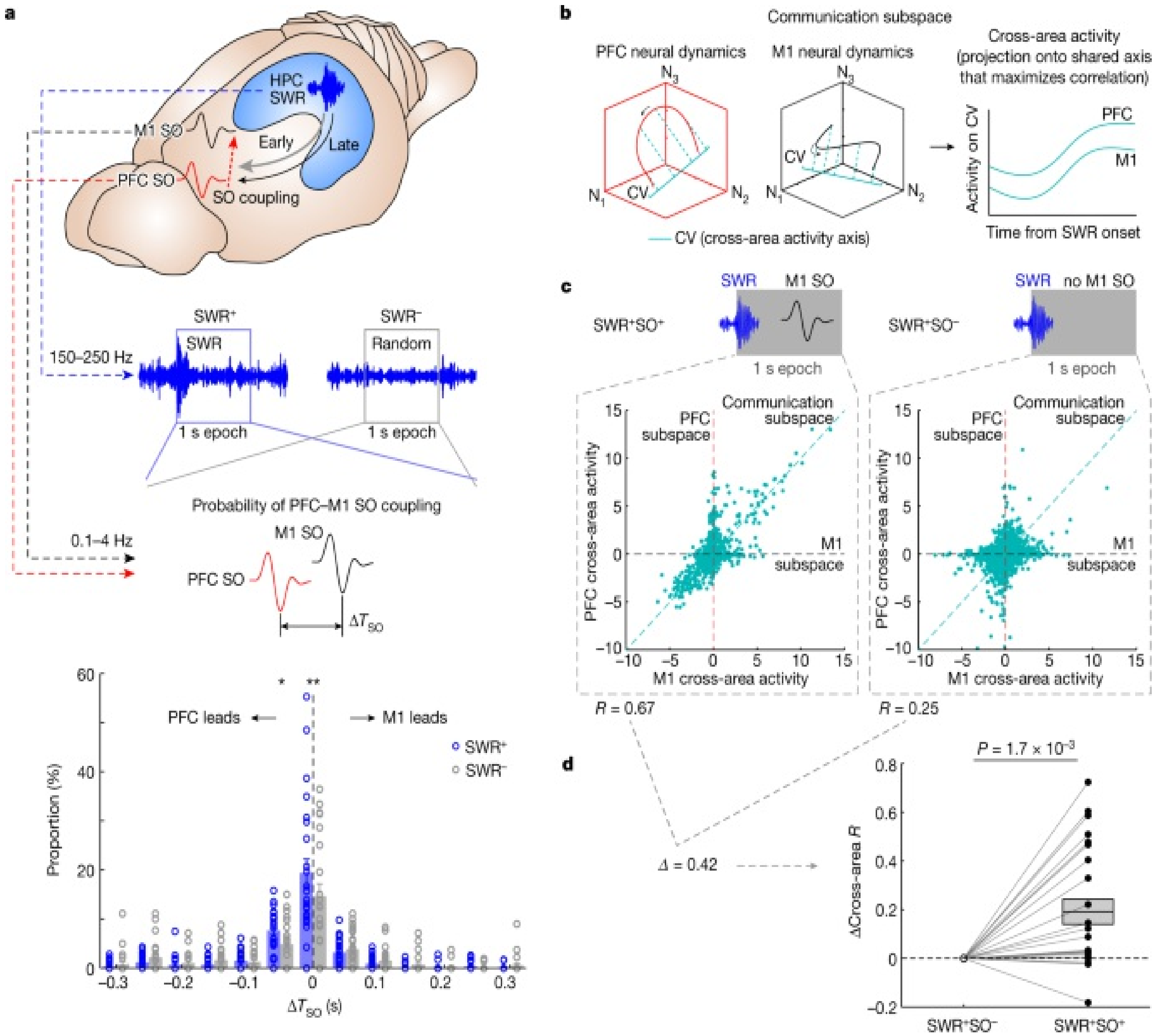
consolidation, Fig. 3d,e). Notably, the period of highly variable success rate (Fig. 1f) also appears to be closely related to this transition. Together, these findings raise the possibility that the transition from exploration—associated with variable performance and a low fidelity—to consolidation is temporally linked to changes in PFC–M1 coupling.

We then repeated the same analyses for PFC (Fig. 3d,e). We found that manifold consolidation in PFC occurred in a delayed fashion relative to M1; this appears to be closely linked to the delayed disengagement of PFC SO–SWR coordination. Thus, changes in ΔSO–SWR coupling appear to follow the onset of neural trajectory stabilization in each area, PFC and M1.

SWRs predict increased PFC–M1 communication

The current model of systems consolidation posits that hippocampal SWRs serve to enhance coordination across cortical networks[9,11,21]. We, therefore, asked whether the presence of SWRs predicts increased cross-area communications between PFC and M1. We focused on the probability of PFC–M1 SO coupling during SWRs (Fig. 4a). We measured the temporal lags between M1 SOs up-states and the nearest PFC SOs up-states (ΔT_{SO}). During early sessions (days 1–4), we found that the distribution of ΔT_{SO} indicated stronger PFC–M1 SO coupling during a 1 s epoch following an SWR (SWR^{+}) compared with a randomly selected 1 s epoch of NREMS (SWR^{-}) (no significant difference in the distributions during the late sessions; Extended Data Fig. 7a).

Fig. 4: Higher PFC–M1 communication is associated with SWR-linked events.



a, Comparison in distributions of PFC SOs from M1 SOs (ΔT_{SO}) between the SWR$^+$ (1 s epochs following SWRs onsets) and the SWR$^-$ condition (randomly selected 1 s NREMS epochs) in the early period (days 1–4). *P = 0.012 and **$P = 7.0 \times 10^{-3}$: significant difference in each 50 ms bin between two conditions using two-sided paired t-test (n = 24 sessions in 6 rats). Mean ± s.e.m. **b**, Schematic diagram showing the communication subspace (cross-area activity axes/canonical variable (CV)) in PFC and M1. Neural dynamics (three neurons' spike activity, $N_{1–3}$) in each area are projected onto the respective CV that maximizes correlation between two areas, that is, cross-area activity showing correlated time course between PFC and M1. **c**,**d**, PFC and M1 cross-area activities during the SWR$^+$SO$^+$ and the SWR$^+$SO$^-$ in the

early period of days 1–4 (**c**). There was stronger correlation between PFC and M1 cross-area activities during the SWR^+SO^+ compared with the SWR^+SO^-, which is shown as the difference between the two conditions (ΔCross-area *R* in **d**; $n = 24$ sessions in 6 rats, two-sided paired *t*-test, $P = 1.7 \times 10^{-3}$). Mean ± s.e.m.

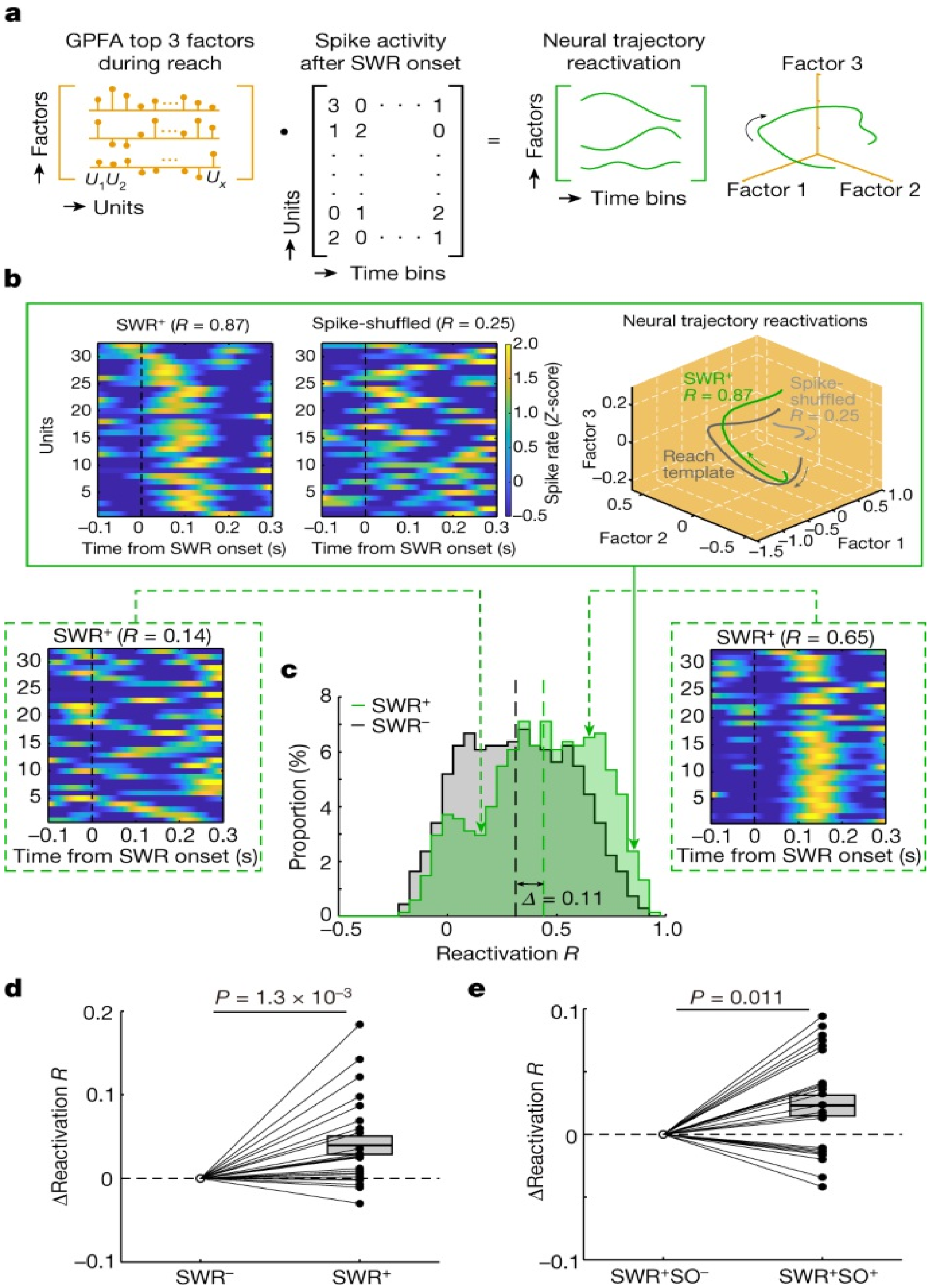
Next, we quantified changes in 'communication subspace'[45] dynamics between PFC and M1 using canonical correlation analysis (CCA)[40] (Fig. 4b). The correlation values of projected data allow us to assess whether there is increased correlated activity across the areas. Cross-area *R* between PFC and M1 was measured during 1 s periods after SWRs (spike peri-event time histogram (PETH) in Extended Data Fig. 7b), and it showed learning-related increases (Extended Data Fig. 7c). Furthermore, cross-area *R* was examined for either SWRs with an SO within 1 s (SWR^+SO^+) or SWRs without a following SO (SWR^+SO^-) (Fig. 4c). The mean cross-area *R* was significantly higher during the SWR^+SO^+ compared with the SWR^+SO^- (SWR^+SO^+, $n =$ 24 sessions in 6 rats: 0.46 ± 0.061 versus SWR^+SO^-, $n = 24$ sessions in 6 rats: 0.26 ± 0.036, two-sided paired *t*-test, $t_{23} = -3.72$, $P = 1.7 \times 10^{-3}$; shown as ΔCross-area *R* in Fig. 4d). Notably, these changes in cross-area communication during SWR^+ and SWR^+SO^+ were not associated with changes in coupling to spindles (Extended Data Fig. 7d). These analyses support the notion that SWRs are linked to significantly elevated PFC–M1 cross-area communication during SO coupling.

M1 reactivation of population dynamics during SWRs

We then asked whether SWRs contained reactivations of task-relevant neuronal activity patterns. Using a computational approach of sleep-associated 'reactivation' of awake ensemble patterns[18,21,46], we examined whether SWRs demonstrate increased strength of ensemble reactivations in M1. We used the top three GPFA factors calculated from the population activity during the awake performance (Fig. 3) to identify NREMS reactivations (Methods and Fig. 5a). We used a single fixed window which demonstrated the highest similarity to the awake template to measure reactivation strength (approximately 200 ms after SWR onset; Extended Data Fig. 8a–c and Methods). We then computed a 'reactivation *R*' between each

reactivation event and the reach template (Fig. 5b; time courses over learning in Extended Data Fig. 8d). When using templates built on spike-shuffled data, we did not find a significant change (Extended Data Fig. 8d). We calculated reactivation *R* values for all SWRs; the mean reactivation *R* value in a single session was chosen to compare SWR^{+} versus SWR^{-} (shown as ΔReactivation *R* in Fig. 5c).

Fig. 5: SWRs with a following SO enhance motor memory reactivations in M1.


a
GPFA top 3 factors during reach
Factors
Units
Spike activity after SWR onset
Units
Time bins
Neural trajectory reactivation
Factors
Time bins
Factor 3
Factor 1
Factor 2
b
SWR⁺ (R = 0.87)
Spike-shuffled (R = 0.25)
Spike rate (Z-score)
Units
Time from SWR onset (s)
Neural trajectory reactivations
SWR⁺ R = 0.87
Spike-shuffled R = 0.25
Reach template
Factor 3
Factor 2
Factor 1
SWR⁺ (R = 0.14)
SWR⁺ (R = 0.65)
c
SWR⁺
SWR⁻
Proportion (%)
Δ = 0.11
Reactivation R
d
P = 1.3 × 10⁻³
ΔReactivation R
SWR⁻
SWR⁺
e
P = 0.011
ΔReactivation R
SWR⁺SO⁻
SWR⁺SO⁺

a, Schematic diagram showing computation of M1 neural trajectory reactivations. Neural ensemble activity during the SWRs period was projected onto the top three GPFA factors defined during the awake experience in Fig. 3; that is, defined as a neural trajectory reactivation. See Methods for the details. **b**, Three examples of neural trajectory reactivations, that is, spike ensemble activity during an SWR. The SWR^+ showed higher correlation to reach the template compared with the SWR-shuffled condition (box with solid line); reactivation *R* was measured comparing a neural trajectory reactivation during sleep with the reach template during task. **c**, Distributions of reactivation *R* values in an example session. We focused on the difference in the mean of reactivation *R* values (vertical dashed lines) between the two comparing conditions (Δ = the value in SWR^+ − the value in SWR.ind.; that is, 'ΔReactivation *R*' in **d** and **e**). Significant difference in two distributions (two-sided two-sample Kolmogorov–Smirnov test, $k = 0.19$, $P = 4.4 \times 10^{-4}$). **d**, Comparisons of the mean of reactivation *R* values between the SWR^+ versus the SWR^- condition ($n = 24$ sessions in 6 rats, two-sided paired *t*-test, $P = 1.3 \times 10^{-3}$). Shown is the difference between the two conditions (ΔReactivation *R*; same as **e**). Mean ± s.e.m. **e**, Comparisons of the mean of reactivation *R* values between the SWR^+SO^+ versus the SWR^+SO^- ($n = 24$ sessions in 6 rats, two-sided paired *t*-test, $P = 0.011$). Mean ± s.e.m.

Mean reactivation *R* values in post-training sleep were significantly higher during the SWR^+ compared with the SWR^- condition during early sessions, days 1–4 ($n = 24$ sessions in 6 rats, SWR^+: 0.38 ± 0.020 versus SWR^-: 0.34 ± 0.022, two-sided paired *t*-test, $t_{23} = -3.65$, $P = 1.3 \times 10^{-3}$; Fig. 5d and Extended Data Fig. 8e for the late sessions, days 10–13). We also conducted two analyses to account for potential changes during NREMS SWRs and coupled SOs. First, the mean firing rate for SWR and SO epochs early and late did not change significantly (Extended Data Fig. 7b). The mean firing rate after SWRs (200 ms window) was not different for early versus late sessions (M1, days 1–4, $n = 24$ sessions in 6 rats: 8.03 ± 0.89 Hz versus days 10–13, $n = 24$ sessions in 6 rats: 8.89 ± 1.10 Hz, LME with two-sided *t*-test, $t_{46} = 1.16$, $P = 0.25$; PFC, days 1–4, $n = 24$ sessions in 6 rats: 8.57 ± 1.04 Hz versus days 10–13, $n = 24$ sessions in 6 rats: 7.02 ± 1.09 Hz, LME with two-

sided *t*-test, $t_{46} = -1.49$, $P = 0.14$). Second, we also shuffled the spiking data within a single SWR spiking epoch (Fig. 5b and Extended Data Fig. 8d). This preserved the mean firing rate within SWR epochs but altered the temporal structure. Using this approach, we did not find evidence for increased reactivation strength after shuffling. Lastly, the mean firing rate in −0.75 to 0.75 s from SO up-states for early sessions was not different from late sessions (M1, days 1–4, $n = 24$ sessions in 6 rats: 7.78 ± 0.85 Hz versus days 10–13, $n = 24$ sessions in 6 rats: 8.74 ± 1.11 Hz, LME with two-sided *t*-test, $t_{46} = 1.29$, $P = 0.20$; PFC, days 1–4, $n = 24$ sessions in 6 rats: 8.23 ± 1.02 Hz versus days 10–13, $n = 24$ sessions in 6 rats: 6.62 ± 1.05 Hz, LME with two-sided *t*-test, $t_{46} = -1.57$, $P = 0.12$).

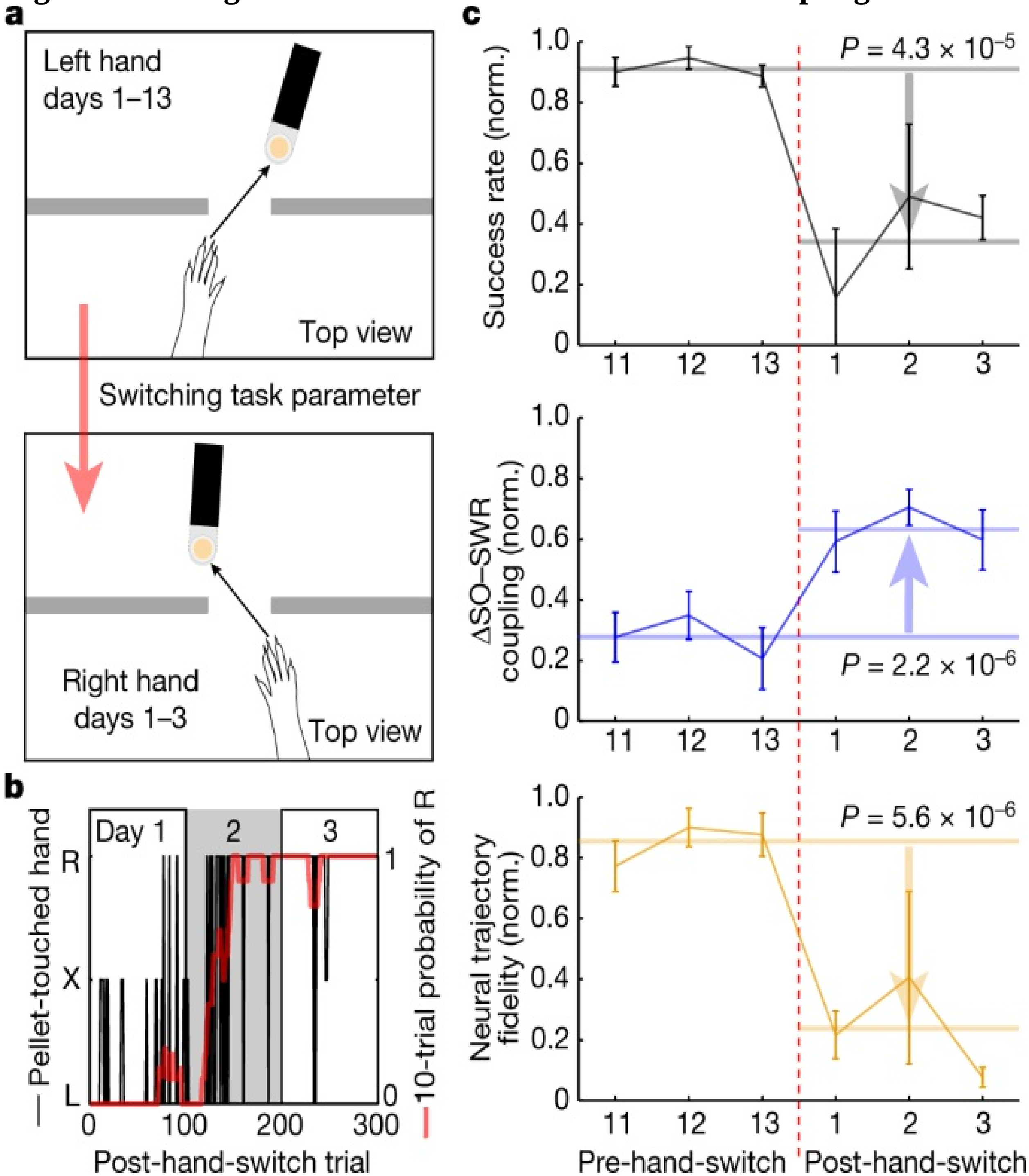
We also found that post-training sleep reactivations were greater than those for pre-training sleep for early sessions; this was not the case for late sessions (Extended Data Fig. 8g,h shows two different comparisons between pre-training versus post-training sleep). This is consistent with past studies which show that reactivations are greater in the post-training sleep[19,47]. Moreover, the mean reactivation *R* values in post-training sleep were significantly higher during the SWR^+SO^+ compared with the SWR^+SO^- ($n = 24$ sessions in 6 rats, SWR^+SO^+: 0.38 ± 0.021 versus SWR^+SO^-: 0.35 ± 0.023, two-sided paired *t*-test, $t_{23} = -2.78$, $P = 0.011$; Fig. 5e and Extended Data Fig. 8f for the late sessions). Together, these findings indicate that stronger M1 reactivations of awake task patterns are highly associated with the presence of SWRs in the early training period.

Re-engaging hippocampal–cortical interactions

Systems consolidation is usually presented as a transfer to cortex with subsequent hippocampal independence. We wondered whether there was a possibility of continuous dialogue. We thus investigated how alterations in task parameters that require new exploration (in this case, need to switch hands) change hippocampal–cortical dialogue. After long-term training using the left hand, in five of the six rats, we changed the pellet location such that the right hand was now required (Fig. 6a). The box and the contextual cues that predicted door opening and task start remained constant. As these animals were overtrained, early exploration was associated with left hand

movements; they did not switch to the right hand until session end (Fig. 6b).

Fig. 6: Switching reach-hand increases M1 SO–SWR coupling.



a, Cartoon of hand-switch experiment. The pellet was located at one side of the wall opening to train animals to use only the dominant hand throughout

approximately 13 d. Then, animals were trained to use the opposite hand by locating the pellet on the other side of the wall opening. **b**, Profile of hand use over time (black; L, left; R, right; X, no try) during the post-hand-switch sessions in an example animal. It shows the 10-trial probability of right-hand use (red). **c**, Post-hand-switch changes in pellet retrieval success rate (top), ΔSO–SWR coupling in M1 (middle) and neural trajectory fidelity in M1 (bottom). Horizontal lines indicate the mean during days 11–13 pre-hand-switch (long, $n = 15$ sessions in 5 rats) or during days 1–3 post-hand-switch (short, $n = 12$ sessions in 5 rats); there was a significant increase or decrease after hand-switch (comparison of the session means using LME with two-sided t-test, success rate: $t_{25} = -4.81$, $P = 4.3 \times 10^{-5}$, ΔSO–SWR coupling: $t_{25} = 6.11$, $P = 2.2 \times 10^{-6}$, neural trajectory fidelity: $t_{25} = -5.74$, $P = 5.6 \times 10^{-6}$). These metrics after hand-switch were normalized relative to the maximum-to-minimum before the hand-switch; this allows us to directly compare the changes after hand-switch compared with the above results. Mean ± s.e.m.

As expected, the success rate significantly decreased (pre-switch, $n = 15$ sessions in 5 rats: 0.91 ± 0.023 versus post-switch, $n = 12$ sessions in 5 rats: 0.34 ± 0.13, LME with two-sided t-test, $t_{25} = -4.94$, $P = 4.3 \times 10^{-5}$; normalized to the pre-switch maximum and minimum in Fig. 6c). Notably, in the following NREMS, ΔSO–SWR coupling in M1 was elevated (pre-switch, $n = 15$ sessions in 5 rats: 0.27 ± 0.050 versus post-switch, $n = 12$ sessions in 5 rats: 0.64 ± 0.052, LME with two-sided t-test, $t_{25} = 6.11$, $P = 2.2 \times 10^{-6}$). We also found that this elevated ΔSO–SWR coupling was associated with increased variability in GPFA neural trajectories in M1, that is, lower fidelity (pre-switch, $n = 15$ sessions in 5 rats: 0.85 ± 0.042 versus post-switch, $n = 12$ sessions in 5 rats: 0.27 ± 0.12, LME with two-sided t-test, $t_{25} = -5.74$, $P = 5.6 \times 10^{-6}$). Thus, our results suggest that, even though the spatial cues in the box and contextual task cues were constant, motor exploration, even after systems consolidation, appeared to alter hippocampal–cortical dialogue during NREMS.

Discussion

Our results using skill learning provide insights into the precise spatiotemporal coordination of multiple areas throughout systems

consolidation. Our observed task-specific reactivations during early manifold exploration suggest a close link between phase-locking of cortical SOs, hippocampal SWRs and cortical reactivations. These results also suggest that hippocampal SWRs may represent a preserved mechanism of consolidation for both declarative and procedural memories.

Although skill learning was initially considered to occur independently of the hippocampus[24,35], recent work indicates that the hippocampus may be recruited early[25,26]. Our results further show how temporally precise feedback between hippocampus and cortex might enable two-stage consolidation and stabilization of M1 neural trajectories. We wondered how we could reconcile this with classic studies of patients with medial temporal lobe lesions who demonstrate skill learning despite lack of explicit recall[48]. Although such patients could clearly learn new skills, their performance levels were somewhat worse than age-matched controls[49]. Thus, although medial temporal lobe-dependent processing may not be completely necessary and remaining systems can partially compensate, it seems increasingly possible that the hippocampus can facilitate consolidation of skills.

We investigated how the hippocampus and PFC might support motor learning. In most experimental skill learning tasks (and in real-world settings), there are contextual cues that inform timing and action selection. In our reaching task, animals must learn to associate allocentric spatial aspects of the box (for example, door and pellet location) with a reaching action. There is also a sound to predict door opening. It is possible that PFC activity patterns[50] support M1 in predicting task rules and the timing of reach. When considering these components of ‘motor learning’, there may also be overlap with instrumental learning (that is, association of actions and outcomes), which may involve hippocampal processing[51]. Substantial future work is required to uncover the precise causal roles of these regions in skill learning and adaptation.

We found that changes in the variability of task activity were temporally associated with stepwise changes in cross-area coupling. Feedback based on task performance and manifold variability may be a key driver to cease hippocampal learning-related engagement and shifts in coordination. Our results also indicate that dialogue across these systems does not end after

systems consolidation. In previous studies, reset of pre-existing connectivity stabilized by consolidation was found to facilitate new learning[52,53]. It is quite possible that 'unlearning' and entering an exploratory state is supported by sleep processing.

The original notion of systems consolidation did not specify precisely what integration into cortex entailed. After apparent hippocampal disengagement, we observed steady increases in fidelity. It is possible that our observed stages are related to fast and slow learning[54,55]. Steady increases in fidelity could be due to greater engagement of the striatum[56], particularly during NREMS[33]. Moreover, although the mechanisms driving shifts from stage I to II are unclear, reward-based processing and dopamine dynamics may be important[57]. Interestingly, SWR-associated reactivations are also found in areas linked to reward processing[58]. How consolidation interacts with reward systems is an intriguing line of inquiry.

Methods

Animals/surgery

Experiments were approved by the Institutional Animal Care and Use Committee at the San Francisco VA Medical Center. We used a total of six adult Long-Evans male rats (300–400 g; Charles River Laboratories). No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications[18,19,20]. Animals were kept under controlled temperature and a cycle with 12 h light and 12 h dark (lights on at 6:00). Surgeries were performed under isoflurane (1–3%) anaesthesia and body temperature was maintained at 37 °C with a heating pad. Atropine sulfate was also administered intraperitoneally before anaesthesia (0.02 mg per kg of body weight). We implanted 64-channel multishank microwire arrays for recording LFP/spike activity (Fig. 1b). One strand of 32-channel arrays was lowered down to 1.4–1.8 mm in layer 5 of the M1 (histology confirmation in Extended Data Fig. 1a) in the forelimb area centred at 3.0 mm lateral and 0.5 mm anterior from the bregma. The other strand of 32-channel arrays was lowered to 3.5–4.0 mm of PFC centred at 1.0 mm lateral and approximately 3.5–4.0 mm anterior from the bregma. In

addition, we implanted 32-channel single-shank linear electrode arrays for recording LFP in the CA1 area of dorsal hippocampus; the tip of electrode arrays was lowered with the angle of 20 degrees to target the coordination of −3.6 mm in AP, 3.0 mm in ML and 4.0 mm in DV from the bregma. Then we adjusted the tip position approximately 200 μm to ensure the best spike recordings in potential CA1 areas (specific three channels ranging 250 μm longitudinally). The reference and ground wires were wrapped around a screw inserted in the midline over the cerebellum. The final localization of depth was based on the quality of recordings across the array at the time of implantation. The post-operative recovery regimen included administration of buprenorphine at 0.02 mg kg^{-1} and meloxicam at 0.2 mg kg^{-1}. Dexamethasone at 0.5 mg kg^{-1} and trimethoprim sulfadiazine at 15 mg kg^{-1} were also administered postoperatively for 5 d. All animals were allowed to recover for at least 7 d with the same regimen as described above before the start of experiments. The food restriction was conducted from 2 d before the start of the experiment until the completion of the experiment; half of the standard amount of food (50 g kg^{-1}) was given to an animal during this period. During behavioural assessments, we monitored the animals and ensured that their body weights did not drop below 90% of initial weight. Data collection and analysis were not performed blind to the conditions of the experiments.

Electrophysiology

We conducted a.c.-coupled recordings and recorded extracellular neural activity using double-stranded 64-channel microwire electrode arrays (33 μm diameter, 250 μm spacing, 2.5 mm length and 4 rows in a single 32-channel strand; 33 μm diameter, 250 μm spacing, 5.0 mm length and 4 rows in the other single 32-channel strand; standard polyimide-coated tungsten microwire arrays from Tucker-Davis Technologies (TDT)) and 32-channel linear electrode arrays (15 μm thickness, 10 mm length, 50 μm spacing and single shank; standard silicone probes from NeuroNexus). All electrode arrays showed similar quality of LFP (for example, LFP amplitude and noise level). We recorded spike and LFP activity using a 128-channel RZ2 bioamp processor (TDT) and 128-channel neurodigitizer (digital PZ5).

Spike data were sampled at 24,414 Hz and LFP data at 1,018 Hz. ZIF-clip-based headstages (TDT) for 64/32-channel electrode arrays with a unity gain and high impedance (approximately 1 G) were used. Only clearly identifiable units with good waveforms and a high signal-to-noise ratio were used. The remaining neural data (for example, filtered LFP) were recorded for offline analyses at 1,018 Hz. Behaviour-related timestamps (that is, trial onset and trial completion) were sent to the RZ2 analog input channel using a digital board and then used to synchronize to neural data in the offline analyses. The LFP was analysed after removing obvious artefacts (>10 s.d.) and excluding bad channels. We used all M1 channels excluding bad channels covering the forelimb-related area (M1 channel count: 23–32 channels, 29.7 ± 2.2, mean ± s.d.). Although the coverage of 32-channel microwires in PFC was large, we did not make any selection of particular channels to include specific PFC areas (for example, infralimbic cortical area known to be important for rule learning), because we focused on more the global activity (that is, SOs) in PFC rather than spike ensemble activity. Thus, we also used all good channels in PFC (PFC channel count: 18–32 channels, 28.3 ± 3.2, mean ± s.d.). However, we located PFC channels to collect spikes and LFP in the medial PFC including the infralimbic area.

Behaviour

After recovery from probe implantations, animals were typically handled for several days (3–5 d) before the start of experimental sessions, that is, ‘day 1’ of long-term recording in which rats showed a successful trial in motor reach-to-grasp task. Animals were acclimated to a custom plexiglass behavioural box during this period without motor training. The box was equipped with a big door at one end for the animal entry and a tiny door at the side for the motor training. After the initial acclimation period, animals were trained for a reach-to-grasp single-pellet task. Typically, animals showed a successful trial at days 1–3 after the start of motor training; the sessions before the day of the first successful trial were not considered as long-term learning sessions shown as days 1–13 in Fig. 1a. The first reach training session (day 1 of long-term motor learning in Fig. 1a) was during days 11–13 (12.5 ± 0.84, mean ± s.d.) after probe implantations; probe implantations were performed contralateral to the preferred hand. Animals were trained to a plateau level of

performance in a reach-to-grasp single-pellet task during 12–13 d (12 d in one animal and 13 d for the other five animals). In a single session of the motor task, pre-training sleep, reach training and post-training sleep were monitored in sequence. The reach-to-grasp task has been used as a sensitive measure of motor function; it requires reaching, grasping and retrieving a single pellet located at a distance outside of the behavioural box[13,14,59]. The pellet was located at one side of the wall opening to train animals to use only the dominant hand (Fig. 6a).

Across animals, we compared relative pellet retrieval success rate in the reach-to-grasp task, using normalized metrics relative to the maximum and minimum success rate; that is, the normalized metric ranged between one and zero corresponding to the maximum and the minimum of the absolute metric, respectively. However, in the hand-switch experiment (Fig. 6), the reach performance after the hand-switch was normalized relative to the maximum-to-minimum before the hand-switch; this was to directly compare the changes after hand-switch with the levels before hand-switch. We monitored electrophysiology, that is, LFP/spike only during the pre-training sleep, reach training and post-training sleep, not during 24 h of a single day. This typically totalled a period of 4–5 h per day starting at 8:30 to 9:30 in the behavioural box. After completing motor tasks and sleep sessions in the behavioural box, animals were placed in the home cage without electrophysiology monitoring.

For the behavioural task, we used an automated reach-box, controlled by custom MATLAB scripts and an Arduino microcontroller. This setup for the reach-to-grasp task required minimal user intervention, as described previously[60]. Each trial consisted of a pellet dispensed on the pellet tray, followed by a beep indicating that the trial was beginning; this was followed by the door opening. Animals then had to reach their arm out, grasp and retrieve the pellet. A real-time ‘pellet detector’ using an infrared detector centred over the pellet was used to determine when the pellet was moved, which indicated that the trial was over and the door was closed. All trials were captured by video, which was synced with the electrophysiology data using the Arduino digital output. The video frame rate was 75 Hz. The reach performance (the number of accurate pellet retrievals/the total number of trials × 100%) was determined manually from a recorded video. The reach performance was used as a measure of motor learning and memory

consolidation across sessions. We also examined task errors, that is, the change in success rate (Extended Data Fig. 2e and Fig. 1f). This was calculated by subtracting the 2 d history success rate from the current day success rate; the 2 d before day 1 was set to zero. For example, the mean success rate of day 5 was subtracted by the mean success rate of days 3–4. After computing the task errors using the absolute success rate, this was normalized in the same manner as for the absolute success rate.

Identification of NREMS oscillations

LFP activity was recorded using 64-channel microwire electrode arrays and 32-channel linear electrodes arrays (see above). The LFP was analysed after removing obvious artefacts (>10 s.d.) and excluding bad channels. Identification of NREMS epochs was performed by classification based on power spectral density of the LFP. This study focused on the NREMS epochs detected using M1 LFP. However, we further conducted NREMS detections using all three areas (that is, PFC, M1 and hippocampal CA1). The total duration of NREMS was not significantly different across the recorded areas (Extended Data Fig. 1c). Moreover, cutting out the last NREMS epoch did not affect the transition trend in ΔSO–SWR coupling (Extended Data Fig. 1d). In detail, regarding NREMS epoch detection, the LFP trace was segmented into non-overlapping 6 s epochs. In each epoch, the power spectral density was computed and averaged over the slow-wave frequency band (0.1–4 Hz, also called the delta band) and gamma frequency bands (30–60 Hz). Then a *k*-means classifier was used to classify epochs into two clusters, NREMS and rapid eye movement sleep (REMS)/awake; REMS and awake were not classified and NREMS was focused on in this study. Sleep epochs less than 30 s were excluded from NREMS epochs. The identified NREMS durations were not different between the early period and late period of motor learning. The identified NREMS epochs were verified by visual assessment of the LFP activity. During the NREMS period with high delta power (0.1–4 Hz), strong down- and up-states dominated. Thus, we assessed whether our detected NREMS epochs contained a high-amplitude and slow LFP fluctuation distinguished from a low-amplitude and high-frequency LFP during the awake period. Moreover, we visually assessed whether there were substantially many wrong detections of NREMS epoch. In other words, we

assessed if a high-amplitude and slow-wave LFP epoch was not included in the detected NREMS epochs. These power-based sleep detections showed a close match to the video-based detections[18]; the number of pixels that changed intensity frame to frame in each pair of consecutive frames was computed from a recorded video (1 Hz frame rate using Microsoft LifeCam Cinema Webcam) during the sleep block; these values were then integrated over an epoch of 40 s. If the integrated value was higher than a threshold, that epoch was identified as sleep; the threshold was chosen by comparing detection results and visual assessment of the recorded video.

In offline analysis, SOs and spindles in the deep layer of PFC and M1 were detected with the algorithm used in previous studies[18,20,33,61]. The recorded layer of PFC and M1 was confirmed with histology (Extended Data Fig. 1a). We also confirmed that the waveform and spike activity during the detected SOs were comparable to the previous publications[18,20,38,62]. The LFP average across all recording channels excluding bad channels was filtered in the delta band (0.1–4 Hz) through two independent filterings: the high-pass Butterworth filter (second order, zero phase-shifted, with a cutoff at 0.1 Hz) was applied and then followed by the low-pass Butterworth filter (fifth order, zero phase-shifted, with a cutoff at 4 Hz). The individual orders of the high-pass and low-pass filters were estimated through a conventional minimum-order design; it required to meet maximum passband ripple of 3 dB and minimum stopband (presumed 0.02 Hz for high-pass filter and 6 Hz for low-pass filter) attenuation of 15 dB. Next, all positive-to-negative zero crossings during NREMS were identified, along with the previous peaks, the following troughs and the surrounding negative-to-positive zero crossings. Then the positive threshold (the top 15 percentile of the peaks) and the negative threshold (the bottom 40 percentile of the troughs) were, respectively, defined for the down-states (that is, silence of neural spiking) and up-states (that is, increased spiking); in Fig. 1b, horizontal dashed lines indicate the thresholds detecting SO up-/down-states. Each identified wave was considered an SO if the trough was lower than the negative threshold (that is, up-state), the peak preceding that up-state was higher than the positive threshold (that is, down-state) and the duration between the peak and the trough was between 150 ms and 500 ms (Fig. 1b).

For spindles detection in PFC and M1, the LFP was first *Z*-scored in each

channel and averaged across all good channels as for the SOs detections. The LFP average was filtered in a spindle band (10–15 Hz) through two independent zero phase-shifted filterings (Extended Data Fig. 5a): the high-pass Butterworth filter (sixth order, zero phase-shifted, with a cutoff at 10 Hz) was applied and then followed by the low-pass Butterworth filter (eighth order, zero phase-shifted, with a cutoff at 15 Hz). The individual orders of the high-pass and low-pass filters were estimated through a conventional minimum-order design; it required to meet maximum passband ripple of 3 dB, and minimum stopband (presumed 7 Hz for high-pass filter and 19 Hz for low-pass filter) attenuation of 15 dB. We computed a smoothed envelope of this signal, the magnitude of the Hilbert transforms with convolving by a Gaussian window (200 ms). Next, we determined two thresholds for spindle detection based on the mean (μ) and standard deviation (σ) of the spindle band LFP during NREMS; the upper and lower thresholds were set $\mu + 2.5 \times \sigma$ and $\mu + 1.5 \times \sigma$, respectively; in Extended Data Fig. 5a, horizontal dashed lines indicate the thresholds detecting spindle peak and spindle period, respectively. Epochs in which the spindle power exceeded the upper threshold for at least one sample and in which the spindle power exceeded the lower threshold for at least 500 ms were considered spindles. Each epoch for which the spindle power exceeded the lower threshold was considered the start and stop of the spindle; the duration of each spindle was based on these values as well.

For SWRs detection in hippocampus, the LFP was first *Z*-scored in each channel and averaged across three channels located on the dorsal CA1 area; during the surgery the electrode tip position was adjusted approximately 200 µm to ensure the best spike recordings in potential CA1 areas (specific three channels ranging 250 µm longitudinally). The LFP average was filtered in the ripple band (150–250 Hz) through two independent zero phase-shifted filterings (Fig. 1b); the high-pass Butterworth filter (eighth order, zero phase-shifted, with a cutoff at 150 Hz) was applied and then followed by the low-pass Butterworth filter (tenth order, zero phase-shifted, with a cutoff at 250 Hz). We computed a smoothed envelope of this signal, the magnitude of the Hilbert transforms with convolving by a Gaussian window (20 ms). Next, we determined two thresholds for SWRs detection based on the mean (μ) and standard deviation (σ) of the SWR band LFP during NREMS; the upper and lower thresholds were set $\mu + 4 \times \sigma$ and $\mu + 1 \times \sigma$, respectively; in Fig. 1b, the

horizontal dashed line indicates the lower threshold. Epochs in which the SWR power exceeded the upper threshold for at least one sample and in which the SWR power exceeded the lower threshold for at least 50 ms were considered SWRs. Each epoch for which the SWR power exceeded the lower threshold was considered the onset and end of the SWR; the duration of each SWR was based on these values as well.

Sleep oscillations coupling analyses

We analysed the temporal coupling of SWRs relative to SOs. For the coupling of SWRs to SOs (SO–SWR coupling; Fig. 2a), each SWR was linked to the closest SO. The time difference between the peak of the SWR and the up-state of the linked SO was measured for each detected SWR ($\Delta T_{\mathrm{SO–SWR}}$). If $\Delta T_{\mathrm{SO–SWR}}$ was between −0.75 s and 0.75 s (that is, nesting time window, black box in Fig. 2b and Extended Data Fig. 3a), that SWR event was considered an SO-coupled SWR. To quantitatively assess the changes in the temporal coupling of SWRs to SOs, we specifically measured the following: the time lag of the SWR from the closest SO ($\Delta T_{\mathrm{SO–SWR}}$) was measured for each SWR event and the rate of SWRs for which $\Delta T_{\mathrm{SO–SWR}}$ was within the nesting time window was measured; that is, the number of SO-coupled SWRs/the total number of SWRs × 100% (Fig. 2b,c). To quantify ‘training-related’ change in SO–SWR coupling over days, we measured ‘ΔSO–SWR coupling’ as the difference between post-training sleep and pre-training sleep (Fig. 2d). We also measured the temporal coupling of SOs observed in PFC and M1, focusing on post-training because the temporal coupling of SOs changes with a slower timescale compared with SO–SWR coupling. For the coupling of PFC SOs and M1 SOs (PFC–M1 SO coupling; Fig. 1c), each M1 SO was linked to the closest SO. The time difference between the up-state of M1 SO and the up-state of the linked PFC SO was measured for each detected M1 SO (ΔT_{SO}). If ΔT_{SO} was between −0.2 s and 0.2 s, that M1 SO was considered a coupled SO. Otherwise, M1 SO was considered an uncoupled SO. We then calculated PFC–M1 SO coupling as the number of coupled SOs in PFC and M1 divided by the total number of SOs (coupled + uncoupled) in M1 (Fig. 1c).

The SWR-shuffled condition in Extended Data Fig. 3c represents temporal

coupling that was measured using the intact SOs and temporally disrupted structure of SWRs. The temporal profile of SWRs in a sleep session was disrupted by circular permutation. The measurement of SO–SWR coupling with circular permutation was repeated 1,000 times in each session and the mean of those 1,000 measures was reported. The same shuffling process was applied to the PFC SO-shuffled M1 SO coupling in Extended Data Fig. 2b. In this case, temporal coupling was measured using the intact PFC SOs and temporally disrupted structure of M1 SOs.

The effect of SWRs on the temporal coupling between PFC SO and M1 SO was examined using the same measurement of ΔT_{SO} in Fig. 4a as well as on the motor memory reactivations in Fig. 5d. We examined it during the SWRs (SWR^{+}) and the random events (SWR^{-}). The SWR^{+} condition represents 1 s time epochs following an SWR onset. The SWR^{-} was defined as the 1 s time epoch following an event that was randomly selected during NREMS; the total number of events of the SWR^{-} condition was set to the number of SWRs for the SWR^{+} condition in each same session. We also examined the intercortical communications and the reactivations during the period of SWRs (Figs. 4d and 5e, respectively); specifically, we compared the SWRs with an SO within 1 s ($SWR^{+}SO^{+}$) versus the SWRs with no following SO in 1 s ($SWR^{+}SO^{-}$). For this category, we focused on more M1 SOs rather than PFC SOs. Thus, the $SWR^{+}SO^{+}$ was defined as an SWR followed by an M1 SO within a 1 s period after that SWR onset, whereas the other SWRs without a following M1 SO within 1 s were considered as $SWR^{+}SO^{-}$.

We also analysed the temporal coupling of spindles relative to SOs. For the nesting of spindles to SOs (SO–Spindle coupling; Extended Data Fig. 5b), each spindle was linked to the closest SO. The time difference between the peak of the spindle and the up-state of the linked SO was measured for each detected spindle ($\Delta T_{SO\text{–}Spindle}$). If $\Delta T_{SO\text{–}Spindle}$ was between −0.5 s and 1.0 s (that is, nesting time window), that spindle event was considered an SO-nested spindle. To quantitatively assess the changes in the temporal coupling of spindles to SOs, we specifically measured as the following: time lag of spindle from the closest SO ($\Delta T_{SO\text{–}Spindle}$) was measured for each spindle event and the rate of spindles for which $\Delta T_{SO\text{–}Spindle}$ was within the nesting time window was measured; that is, the number of SO-nested spindles/the

total number of spindles × 100% (Extended Data Fig. 5c,d). We observed 'triple-coupling' (that is, precise coupling of SWR, SO and spindle; Extended Data Fig. 4b): the percentage of SWRs locking with both SO and spindle between −1 s and 1 s from an SWR peak. We also observed 'quadruple-coupling' (that is, precise coupling of SWR, coupled SOs in M1 and PFC, and spindle; Extended Data Fig. 6e): the percentage of SWRs locking with PFC–M1-coupled SOs, and spindle between −1 s and 1 s from an SWR peak.

Spike activity during events

We initially used an online sorting programme (SpikePac, TDT). We then conducted offline sorting using MountainSort for six rats[63]. Briefly, MountainSort is a spike-sorting software that uses an automatic algorithm, which compares clusters of data and identifies single units using one-dimensional projections. After the automatic sorting using MountainSort, a minimal manual merging and rejection of clusters were performed. Only clearly identifiable units along with sessions of a single day, with good waveforms and a high signal-to-noise ratio, were used. There was no significant change in the number of recorded units across long-term motor learning in both M1 and PFC (the number of units per session in M1, early, $n = 24$ sessions in 6 rats: 30.54 ± 1.35 per session versus late, $n = 24$ sessions in 6 rats: 29.00 ± 0.97 per session, LME with two-sided *t*-test, $t_{46} = -1.22$, $P = 0.23$; the number of units per session in PFC, early, $n = 24$ sessions in 6 rats: 25.50 ± 1.61 per session versus late, $n = 24$ sessions in 6 rats: 25.46 ± 1.68 per session, LME with two-sided *t*-test, $t_{46} = -0.026$, $P = 0.98$).

To assess spike activity modulation during awake training epochs and sleep oscillations, we analysed PETH. After spikes were time-locked to event reference times (for example, reach onsets during task or SWR onsets during sleep), the PETH (bin length = 15 ms) was estimated. The PETH during a reach onset was calculated in an approximately 600–1,200 ms window depending on the mean reach-to-pellet-touch duration for each animal (−200 to 400 ms for 4 rats, −200 to 600 ms for 1 rat and −200 to 1,000 ms for 1 rat from a reach onset; see also the 'GPFA' section below). The PETH during an SWR onset was calculated in 195 ms and 1,000 ms windows after the SWR onset, respectively, for the reactivation and the CCA analysis (see the

detailed method of window size selection in the ‘Memory reactivation analyses’, and more details in the ‘CCA’ section below). The PETHs in Fig. 3 and Fig. 5 are shown after normalization in each unit by subtraction of mean and then division by standard deviation in each event window.

The spike-shuffled condition in Extended Data Fig. 6c and Extended Data Fig. 8d was conducted using the circular permutation in an individual unit. Among the recorded unit population, the timing of spike events in the first unit was permutated around an event time in a circular way and then it was repeated along with the recorded units as well as along with the monitored events. The time window for the circular permutation was approximately three times longer than the analysis window described above for the GPFA and the reactivation analysis; circular permutation windows were −600 to 1,000 ms for the GPFA analysis and −200 to 400 ms for the reactivation analysis. The one set of circular permutations in each session was repeated 1,000 times and the mean of those 1,000 measures was reported.

CCA

To identify a shared cross-area subspace (that is, communication subspace between PFC and M1 in Fig. 4), we used CCA which identifies maximally correlated linear combinations between two groups of variables[40]. As illustrated in the cartoon in Fig. 4b, CCA allows us to identify cross-area activity axes (CV) on which projections of local PFC neural dynamics (red) and M1 neural dynamics (black) are maximally correlated during SWRs. CCA was carried out using the MATLAB function ‘canoncorr’. Neural data in PFC and M1 during SWRs were binned at 15 ms, and data during 1 s from SWR onsets were concatenated across events and mean subtracted. Then CCA models were fit to these data in each session. CCA produces as many CVs as the number of neurons in the smaller population (for example, if there are 10 PFC neurons and 15 M1 neurons, then CCA will fit 10 CVs). For evaluating cross-area activities, only the top CV explaining maximally correlated cross-area activity among CVs was used, as this provided a consistent dimensionality across datasets, and a signal with both magnitude and sign. Then we calculated ‘cross-area correlation (R)’: the correlation coefficient between PFC cross-area activity and M1 cross-area activity was computed for SWR periods in Extended Data Fig. 7c as well as for the

periods of $SWR^{+}SO^{+}$ and $SWR^{+}SO^{-}$ in Fig. 4c. The stronger communication is demonstrated as alignments of cross-area activities with 'communication subspace' (diagonal dashed line in Fig. 4c) resulting in a higher cross-area *R* value. On the contrary, the weaker communication or independent activity either in PFC or M1 is demonstrated as alignments with the PFC subspace (vertical dashed line) or M1 subspace (horizontal dashed line) resulting in a lower cross-area *R* value.

GPFA

To characterize single-trial representations of population spiking activity in Fig. 3, we used GPFA[29] to find low-dimensional neural trajectories for each trial. GPFA analyses were carried out for an individual area of PFC or M1 using the MATLAB-based DataHigh (v.1.3)[64]. As shown using the diagram in Fig. 3a, this method allows us to decompose neural population activity ($x = \mu + x^{shared} + x^{private}$; in which μ represents mean spike firing rate computed in each neuron) into shared activity ($x^{shared} = Uz$; in which U and z represent shared factor weight and shared factor activity, respectively) and private activity ($x^{private}$); the top cartoon in Fig. 3a is shown using two neurons (N_1 and N_2). The bottom cartoon illustrates a transition from 'manifold exploration' in early learning periods (for example, days 3 and 4) characterized with highly variable shared activity to 'manifold consolidation or stabilization' in the late learning period (for example, days 9 and 13) characterized with the emergence of a manifold and stable neural trajectories over long days.

GPFA was performed with a time bin of 15 ms and the optimal dimensionality of 6 which was determined by the method of 'leave-neuron-out strategy' proposed in the previous study[29]; the first factor accounts for the largest variance in the neural population activity during task training. For each session, we concatenated binned spike trains for each neuron across trials. This concatenated spike train was *Z*-transformed to account for neurons with high firing rates. Then the *Z*-transformed spike trains were placed into a two-dimensional matrix organized by neurons (x) and time (y, the number of time bins). From this spike count matrix, the shared factor weight (also called GFPA factors) and shared factor activity (projections of the *Z*-transformed

spike trains onto the GPFA factors) were computed. The top three factors were used for the analyses; they accounted for >85% of shared variance explained in respective M1 and PFC in each session (example trajectories in M1 in an animal, Extended Data Fig. 6a,b). We confirmed that shared-over-total variance, that is, shared variance divided by the sum of shared and private variance, showed a robust increase during learning (Extended Data Fig. 6c). However, the shared-over-total variance with no significant change in the spike-shuffled condition (see above for the details about the circular permutation) supports the idea that the temporal pattern of neural ensemble resulted in the robust learning in neural dynamics and not the simple change in population firing rate. We recalculated the low-dimensional manifold (that is, GPFA factors) every session because recorded units were not able to be held across days. To compensate for possible variations of low-dimensional manifolds, the estimated manifolds were aligned to the average manifold of the final 3 d by using Procrustes alignment (MATLAB function 'procrustes')[41,65].

We examined the consistency and variance of neural trajectories over long-term task learning. First, in an individual animal, we calculated 'GPFA correlation': the correlation coefficient of a single-trial neural trajectory with the optimal trajectory calculated by averaging the neural trajectories of the three sessions in which the animal showed the best performance over the 13 d task training. It provided a robust measure of neural optimality for a successful task as the mean and standard deviation of this measure increased and decreased, respectively, during the long-term learning (Extended Data Fig. 6d). To quantify the optimality as well as stabilization of neural trajectories, we then measured 'neural trajectory fidelity', that is, signal-to-noise ratio[42,43,44], of GPFA correlations across trials. This measure was calculated by the mean over the standard deviation of GPFA correlations across trials and compared the level of a desired neural trajectory ('signal') with the level of deviation from the optimal trajectory ('background noise'); it measures both strength and consistency of GPFA correlations across all trials including both successful and failed trials. For this measurement, 'high fidelity' represents low representational variability and stable neural trajectories, and 'low fidelity' represents greater representational variability. We then define low fidelity as 'manifold exploration', because there is large variation in neural trajectories. With practice and changes in sleep

coordination, we see a transition to steady increases in fidelity. We define this as ‘manifold consolidation’.

GPFA was performed specifically for the period near the reach-to-grasp: depending on the mean duration from reach onset to pellet touch in each animal, for M1 population activity it was between −200 ms and 400 ms from reach onset for four animals, between −200 and 600 ms from reach onset for one animal, and between −200 and 1,000 ms from reach onset for one animal. The same computations were conducted for PFC population activity but using the expanded periods, which were expanded by 100 ms both at the beginning and the end of each epoch for M1. Specifically, for the GPFA correlations, we used neural trajectories during the mean reach-to-grasp period of that session. As this duration varied across trials, sessions and animals, we interpolated each trial such that every trial was the same length (100 values) and then calculated correlation coefficients.

Memory reactivation analyses

To characterize ensemble reactivations during sleep, we performed an analysis that compared neural activity patterns in M1 during post-training sleep using a template that was created based on neural activity during reach task training. The previous studies typically used principal component analysis to convert a set of observations of neural activity during the task into a set of values of linearly uncorrelated variables called principal components[18,19,20,21,66]. In this study we used GPFA factors instead of principal components to directly compare the findings with the concept of neural trajectory in Fig. 3; the basic approach is analogous to the previous studies (Fig. 5a). In detail, we used the top three GPFA factors computed using the spike trains during task trials (see above for the details of GPFA computation using awake activities), that is, the manifold or the template space in which a template of ensemble activity evolves during reach training. The binned spike trains (bin length = 15 ms) during a particular event of post-training sleep (for example, SWR and random event) were then projected onto the ‘template space’, that is, GPFA top three factors defined by the neural activity during reach training; this projection was a linear combination of *Z*-scored binned spike trains from post-training sleep with the template space calculated from the neural activity during reach training. This linear

combination has been termed the ‘neural trajectory reactivation’.

As motor memory reactivation is known to be temporally compressed/extended to different degrees relative to an awake pattern[46,67], we looked for reactivations with different window sizes (75–405 ms corresponding to approximately 0.2–2× compression/extension) as well as with various time lags (0–405 ms) from an SWR onset; we explored the combinations of sizes and time lags to find the best reactivation for each SWR onset (Extended Data Fig. 8a). Here time lag was limited so that the end of prediction windows was not out of the 405 ms period from an SWR onset. In each SWR onset of post-training sleep, the combination showing the maximal correlation with the reach template (mean neural trajectory across the conducted task trials) of that session was selected for the final neural trajectory reactivation for that event; the distributions of window sizes and time lags of the final neural trajectory reactivations are shown in Extended Data Fig. 8b,c. As a simple comparison across trials and sessions, we then used the fixed window size to measure neural trajectory reactivations; the most frequent size and time lag were used for all sessions and trials, that is, 195 ms size and zero time lag from the SWRs onsets. The correlation coefficient between a neural trajectory reactivation during this window and the reach template was then termed the ‘reactivation correlation (*R*)’ of that particular neural trajectory reactivation. We confirmed the robust increase over long-term motor learning (Extended Data Fig. 8d). However, the reactivation *R* did not demonstrate significant changes in the spike-shuffled condition. The goal of the shuffling procedure, done per event epoch, was to maintain the mean firing while disrupting the temporal structure across neurons. This supports the idea that the temporal pattern of the neural ensemble resulted in a robust increase in reactivation *R*.

We examined reactivations for all SWRs; the mean/median of reactivations in a single session was chosen to use. We confirmed that the reactivation *R* values during sleep in observed spike activities of a particular condition were significantly greater than in the spike-shuffled condition, that is, the projections of temporally shuffled neural activity for each unit in each event window onto the template space (GPFA top three factors). The mean/median of the reactivation *R* values computed in each session was then used for the comparison between the SWR^+ versus the SWR^- as well as the SWR^+SO^+

versus the SWR^+SO^- condition.

Across-session analyses over long-term learning

Cross-areas coordination and neural population dynamics were analysed during the long-term motor skill learning in Figs. 1d,f, 2c–e and 3d. To examine the transitions across sessions, we used tertiles; either trials during reach training or events during sleep were divided into tertiles per session. We then convolved the 13 d data (39 tertiles in five animals; 12 d with 36 tertiles in one animal) using a Gaussian window of 3 d (9 tertiles); the data were padded by an average during the first 2 d at the start and by an average during the last 2 d at the end before convolution. As mentioned above, in a single animal, the 12 d motor training was performed; thus, in this animal, days 9–12 were analysed for the late period marked 'days 10–13'.

To extend the reliability of two-stage prediction beyond the variability in absolute values across multiple animals, the metric was normalized in each animal; the normalized metric ranged between one and zero corresponding to the maximum and the minimum of the absolute metric, respectively. The convolved and normalized data were then fitted to the Sigmoid function having a characteristic 'S'-shaped curve which allows us to predict two-stage transitions over time with high versus low values. We used a general form of the sigmoid function, in which U and L were the upper and lower boundaries, respectively, x_{mid} was the midpoint parameter (symmetric point of S-shaped curve) and k was the slope parameter. The midpoint parameter of the sigmoid function informs the temporal distinction of two stages. The sigmoid function fittings were conducted focusing on days 3–11 for M1 and days 3–13 for PFC because of the delayed transition of cross-area coupling in PFC compared with M1. This information was used to determine the sharpest changes in ΔSO–SWR coupling and PFC–M1 SO coupling and to compare the sequential interactions from the stepwise increase in PFC–M1 SO coupling to the drop in ΔSO–SWR coupling (Fig. 2e). For the success rate in Extended Data Fig. 2f, the convolved and normalized data in the same way for PFC–M1 SO coupling were fitted to the single-exponential function to estimate the decay over days ($P = 7.5 \times 10^{-14}$).

We also used the piecewise linear regression method to predict two/three linear trends of changes in the PFC–M1 SO coupling (Fig. 1d,e), the changes in success rate (Fig. 1f) and the neural trajectory fidelity (Fig. 3d). The two/three trends were first classified based on the time course of the activity in six rats. Linear slope across points in a moving window was measured using conventional linear regression (MATLAB function ‘regress’). For example, in the neural trajectory fidelity in Fig. 3d, linear slopes were measured in a 5 d moving window from day 1 to day 13 and then were investigated when the 5 d linear slope increased rapidly by comparing with the values of days 1–4; it was on day 6 and day 9 for M1 and PFC, respectively (Extended Data Fig. 6e). We repeated this using different sizes of windows, that is, from 3 d to 7 d, and found that the linear slope significantly increased at day 6 for M1 and day 9 for PFC whichever window size was used. Thus, we focused on the trends of linear changes over time in the two distinct periods, for example, days 1–5 versus days 6–13 for M1 in each animal. For the days 1–5, we used all 15 tertiles to estimate linear slope; however, for the days 6–13, we used the tertiles from day 6 to the maximum-level tertile to focus on the rising phase and excluding the plateau phase at the latest period of motor learning. We repeated the same process for the PFC–M1 SO coupling (Figs. 1d,e) and the changes in success rate (Fig. 1f) but using all data points.

To test the effect of the rate of PFC SOs in the transitions of the PFC–M1 SO coupling, we recalculated the PFC–M1 SO coupling after subsampling the PFC SOs (Extended Data Fig. 2d). For the observed PFC SOs in a single session (for example, 425 events), 100 PFC SOs (this subsample number was fixed across sessions over time) were randomly subsampled, and then the PFC–M1 SO coupling was recalculated. In each session, this was repeated 1,000 times and the mean of those 1,000 measures was reported.

Change-point analysis

Cross-areas coordination (PFC–M1 SO coupling and ΔSO–SWR coupling) was used to estimate two-stage transition time points during the long-term motor learning in Fig. 2f. We also performed ‘change-point analysis’ using the MATLAB function ‘findchangepts’. This finds a change-point at which some statistical property of a signal changes abruptly[68] (single animal example in Extended Data Fig. 3e). The chosen statistic (type of change to detect) was the mean with setting to find a single change-point over 13 d data; the algorithm minimizes the total residual error from the best horizontal level of the mean for each stage. The stage transitions, that is, abrupt change based on PFC–M1 SO coupling and ΔSO–SWR coupling, were at 6.83 ± 0.17 d (mean ± s.e.m.) across the six animals.

Quantification and statistical analyses

The figures show mean ± s.e.m.; if this was not the case, we specifically indicated so. Parametric statistics were generally used in this study (LME, *t*-tests, linear regression, Pearson’s correlation or otherwise stated), and were implemented within MATLAB. A ‘hierarchical nested statistics approach’ of LME (using the MATLAB function ‘fitlme’) was used for the comparison of SWRs rates, SOs rates and NREMS periods, in the main text and in Extended Data Fig. 1e; for the comparisons across days in Fig. 2b–d and Extended Data Figs. 2b, 3b,c, 4b, 5d, 6c, 7c,d and 8d; for the success rate, ΔSO–SWR coupling and neural trajectory fidelity in Fig. 6c; and for the changes in linear slope in Fig. 1e and Extended Data Figs. 2c and 6e. This was done to account for the repeated measures per animal; thus, this statistical approach ensured that the group level statistic accounted for sessions per animal and did not treat them as statistically independent samples. We fit random effects (for example, rats) specified as an intercept for each group and reported fixed effects representing population parameters to compare (for example, pre-hand-switch versus post-hand-switch). Adding random effects to a model recognizes correlations within sample subgroups (for example, rat) and extends the reliability of inferences beyond the variability across multiple rats. The fixed effects were tested for *P* values of the linear regression slope

coefficients associated with two comparing conditions. The random effects and fixed effects parameters used are as follows: Fig. 2c,d and Extended Data Figs. 2b,c, 3c and 6e, random: rat, fixed: day; Figs. 2b,c and 3e and Extended Data Figs. 1e, 4b, 5d, 6c, 7c and 8d, random: rat, fixed: phase; Fig. 6c, random: rat, fixed: hand-switch. In these figures, the mean in each experimental session was used as the response parameter and two categories of the comparing conditions were used as the predictor parameter. For the comparison between two conditions in Extended Data Fig. 7a, we used two-sided two-sampled *t*-test. For the comparison between two paired conditions in Figs. 1e,f, 2b,f, 3e, 4a,d and 5d,e and Extended Data Figs. 3a and 8e–h, we used two-sided paired *t*-test.

We used traditional linear regression or correlation to evaluate the relationship from the motor learning period to the PFC–M1 SO coupling in Fig. 1d, the behavioural performance in Fig. 1f and the neural trajectory fidelity in Fig. 3d, as well as the relationship between behaviour and PFC–M1 SO coupling in Fig. 1g. For the comparison between distributions, we used Kolmogorov–Smirnov test in Fig. 5c. We also used logistic regression to illustrate the two-stage distinction of motor learning in Fig. 2g. Logistic regression requires predesignated parameter values for the predicting conditions, for example, tagging the stage I versus the stage II for each session. To tag the stage I versus stage II, we used the median of rise time (midpoint of sigmoid function fit) of PFC–M1 SO coupling and the drop time (midpoint of sigmoid function fit) of ΔSO–SWR coupling; this has been termed the 'grand-midpoint'. We, thus, tagged the sessions before the grand-midpoint as stage 1 and the sessions after the grand-midpoint as stage II in each animal; the grand-midpoint occurred near days 7–8 across six animals. Then logistic regression was performed with the stage-tagged sessions to predict two distinct stages as a function of ΔSO–SWR coupling and PFC–M1 SO coupling.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data used in this manuscript are available at https://zenodo.org/deposit/7226711.

Code availability

Custom MATLAB codes for the detection of SOs, spindles and SWRs are available at our GitHub repository (https://github.com/kimjack0/Sleep_Oscillation_Detection).

References

1. Buzsáki, G. Two-stage model of memory trace formation: a role for “noisy” brain states. *Neuroscience* **31**, 551–570 (1989).

2. McClelland, J. L., McNaughton, B. L. & O’Reilly, R. C. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* **102**, 419–457 (1995).

3. Kitamura, T. et al. Engrams and circuits crucial for systems consolidation of a memory. *Science* **356**, 73–78 (2017).

4. Klinzing, J. G., Niethard, N. & Born, J. Mechanisms of systems memory consolidation during sleep. *Nat. Neurosci.* **22**, 1598–1610 (2019).

5. Wilson, M. A. & McNaughton, B. L. Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679 (1994).

6. Sawangjit, A. et al. The hippocampus is crucial for forming non-hippocampal long-term memory during sleep. *Nature* **564**, 109–113 (2018).

7. Jadhav, S. P., Kemere, C., German, P. W. & Frank, L. M. Awake hippocampal sharp-wave ripples support spatial memory. *Science* **336**, 1454–1458 (2012).

8. Girardeau, G., Benchenane, K., Wiener, S. I., Buzsaki, G. & Zugaro, M. B. Selective suppression of hippocampal ripples impairs spatial memory. *Nat. Neurosci.* **12**, 1222–1223 (2009).

9. Khodagholy, D., Gelinas, J. N. & Buzsaki, G. Learning-enhanced coupling between ripple oscillations in association cortices and hippocampus. *Science* **358**, 369–372 (2017).

10. Rothschild, G., Eban, E. & Frank, L. M. A cortical-hippocampal-cortical loop of information processing during memory consolidation. *Nat. Neurosci.* **20**, 251–259 (2017).

11. Siapas, A. G. & Wilson, M. A. Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep. *Neuron* **21**, 1123–1128 (1998).

12. Preston, A. R. & Eichenbaum, H. Interplay of hippocampus and prefrontal cortex in memory. *Curr. Biol.* **23**, R764–773 (2013).

13. Ramanathan, D. S. et al. Low-frequency cortical activity is a neuromodulatory target that tracks recovery after stroke. *Nat. Med.* **24**, 1257–1267 (2018).

14. Guo, L., Kondapavulur, S., Lemke, S. M., Won, S. J. & Ganguly, K. Coordinated increase of reliable cortical and striatal ensemble activations during recovery after stroke. *Cell Rep.* **36**, 109370 (2021).

15. Tse, D. et al. Schemas and memory consolidation. *Science* **316**, 76–82 (2007).

16. Nadel, L. & Moscovitch, M. Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr. Opin. Neurobiol.* **7**, 217–227 (1997).

17. Ganguly, K., Khanna, P., Morecraft, R. J. & Lin, D. J. Modulation of neural co-firing to enhance network transmission and improve motor function after stroke. *Neuron* **110**, 2363–2385 (2022).

18. Kim, J., Gulati, T. & Ganguly, K. Competing roles of slow oscillations and delta waves in memory consolidation versus forgetting. *Cell* **179**, 514–526.e3 (2019).

19. Ramanathan, D. S., Gulati, T. & Ganguly, K. Sleep-dependent reactivation of ensembles in motor cortex promotes skill consolidation. *PLoS Biol.* **13**, e1002263 (2015).

20. Gulati, T., Guo, L., Ramanathan, D. S., Bodepudi, A. & Ganguly, K. Neural reactivations during sleep determine network credit assignment. *Nat. Neurosci.* **20**, 1277–1284 (2017).

21. Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I. & Battaglia, F. P. Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nat. Neurosci.* **12**, 919–926 (2009).

22. Genzel, L. & Robertson, E. M. To replay, perchance to consolidate. *PLoS Biol.* **13**, e1002285 (2015).

23. Massimini, M., Huber, R., Ferrarelli, F., Hill, S. & Tononi, G. The sleep slow oscillation as a traveling wave. *J. Neurosci.* **24**, 6862–6870 (2004).

24. Squire, L. R. & Zola, S. M. Structure and function of declarative and nondeclarative memory systems. *Proc. Natl Acad. Sci. USA* **93**, 13515–13522 (1996).

25. Boutin, A. et al. Transient synchronization of hippocampo-striato-thalamo-cortical networks during sleep spindle oscillations induces motor memory consolidation. *Neuroimage* **169**, 419–430 (2018).

26. Albouy, G. et al. Both the hippocampus and striatum are involved in consolidation of motor sequence memory. *Neuron* **58**, 261–272 (2008).

27. Molle, M., Yeshenko, O., Marshall, L., Sara, S. J. & Born, J. Hippocampal sharp wave-ripples linked to slow oscillations in rat slow-wave sleep. *J. Neurophysiol.* **96**, 62–70 (2006).

28. Joo, H. R. & Frank, L. M. The hippocampal sharp wave-ripple in

memory retrieval for immediate use and consolidation. *Nat. Rev. Neurosci.* **19**, 744–757 (2018).

29. Yu, B. M. et al. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J. Neurophysiol.* **102**, 614–635 (2009).

30. Gallego, J. A., Perich, M. G., Miller, L. E. & Solla, S. A. Neural manifolds for the control of movement. *Neuron* **94**, 978–984 (2017).

31. Churchland, M. M. et al. Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).

32. Athalye, V. R., Ganguly, K., Costa, R. M. & Carmena, J. M. Emergence of coordinated neural dynamics underlies neuroprosthetic learning and skillful control. *Neuron* **93**, 955–970.e5 (2017).

33. Lemke, S. M. et al. Coupling between motor cortex and striatum increases during sleep over long-term skill learning. *eLife* **10**, e64303 (2021).

34. Hikosaka, O., Nakamura, K., Sakai, K. & Nakahara, H. Central mechanisms of motor skill learning. *Curr. Opin. Neurobiol.* **12**, 217–222 (2002).

35. Frankland, P. W. & Bontempi, B. The organization of recent and remote memories. *Nat. Rev. Neurosci.* **6**, 119–130 (2005).

36. Born, J., Rasch, B. & Gais, S. Sleep to remember. *Neuroscientist* **12**, 410–424 (2006).

37. Buzsáki, G. Hippocampal sharp wave - ripple: a cognitive biomarker for episodic memory and planning. *Hippocampus* **25**, 1073–1188 (2015).

38. Kim, J. et al. Recovery of consolidation after sleep following stroke-interaction of slow waves, spindles, and GABA. *Cell Rep.* **38**, 110426 (2022).

39. Ganguly, K. & Carmena, J. M. Emergence of a stable cortical map for

neuroprosthetic control. *PLoS Biol.* **7**, e1000153 (2009).

40. Veuthey, T. L., Derosier, K., Kondapavulur, S. & Ganguly, K. Single-trial cross-area neural population dynamics during long-term skill learning. *Nat. Commun.* **11**, 4057 (2020).

41. Degenhart, A. D. et al. Stabilization of a brain-computer interface via the alignment of low-dimensional spaces of neural activity. *Nat. Biomed. Eng.* **4**, 672–685 (2020).

42. Rieke, F., Warland, D., Van Steveninck, R. D. R. & Bialek, W. *Spikes: Exploring the Neural Code* (MIT Press, 1999).

43. Barlow, H. B. & Levick, W. R. Three factors limiting the reliable detection of light by retinal ganglion cells of the cat. *J. Physiol.* **200**, 1–24 (1969).

44. Zohary, E., Shadlen, M. N. & Newsome, W. T. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* **370**, 140–143 (1994).

45. Semedo, J. D., Zandvakili, A., Machens, C. K., Yu, B. M. & Kohn, A. Cortical areas interact through a communication subspace. *Neuron* **102**, 249–259.e4 (2019).

46. Buch, E. R., Claudino, L., Quentin, R., Bonstrup, M. & Cohen, L. G. Consolidation of human skill linked to waking hippocampo-neocortical replay. *Cell Rep.* **35**, 109193 (2021).

47. Eckert, M. J., McNaughton, B. L. & Tatsuno, M. Neural ensemble reactivation in rapid eye movement and slow-wave sleep coordinate with muscle activity to promote rapid motor skill learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190655 (2020).

48. Gabrieli, J. D., Corkin, S., Mickel, S. F. & Growdon, J. H. Intact acquisition and long-term retention of mirror-tracing skill in Alzheimer's disease and in global amnesia. *Behav. Neurosci.* **107**, 899 (1993).

49. McDougle, S. D., Wilterson, S. A., Turk-Browne, N. B. & Taylor, J. A. Revisiting the role of the medial temporal lobe in motor learning. *J. Cogn. Neurosci.* **34**, 532–549 (2022).

50. Narayanan, N. S. & Laubach, M. Top-down control of motor cortex ensembles by dorsomedial prefrontal cortex. *Neuron* **52**, 921–931 (2006).

51. Corbit, L. H. & Balleine, B. W. The role of the hippocampus in instrumental conditioning. *J. Neurosci.* **20**, 4233–4239 (2000).

52. Epp, J. R., Silva Mera, R., Kohler, S., Josselyn, S. A. & Frankland, P. W. Neurogenesis-mediated forgetting minimizes proactive interference. *Nat. Commun.* **7**, 10838 (2016).

53. Park, A. J. et al. Reset of hippocampal-prefrontal circuitry facilitates learning. *Nature* **591**, 615–619 (2021).

54. Kami, A. et al. Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature* **377**, 155–158 (1995).

55. Doyon, J. & Benali, H. Reorganization and plasticity in the adult brain during learning of motor skills. *Curr. Opin. Neurobiol.* **15**, 161–167 (2005).

56. Lemke, S. M., Ramanathan, D. S., Guo, L., Won, S. J. & Ganguly, K. Emergent modular neural control drives coordinated motor actions. *Nat. Neurosci.* **22**, 1122–1131 (2019).

57. Hosp, J. A., Pekanovic, A., Rioult-Pedotti, M. S. & Luft, A. R. Dopaminergic projections from midbrain to primary motor cortex mediate motor skill learning. *J. Neurosci.* **31**, 2481–2487 (2011).

58. Lansink, C. S., Goltstein, P. M., Lankelma, J. V., McNaughton, B. L. & Pennartz, C. M. Hippocampus leads ventral striatum in replay of place-reward information. *PLoS Biol.* **7**, e1000173 (2009).

59. Whishaw, I. Q., Oconnor, W. T. & Dunnett, S. B. The contributions of

motor cortex, nigrostriatal dopamine and caudate-putamen to skilled forelimb use in the rat. *Brain* **109**, 805–843 (1986).

60. Wong, C. C., Ramanathan, D. S., Gulati, T., Won, S. J. & Ganguly, K. An automated behavioral box to assess forelimb function in rats. *J. Neurosci. Methods* **246**, 30–37 (2015).

61. Silversmith, D. B., Lemke, S. M., Egert, D., Berke, J. D. & Ganguly, K. The degree of nesting between spindles and slow oscillations modulates neural synchrony. *J. Neurosci.* **40**, 4673–4684 (2020).

62. Sirota, A. & Buzsáki, G. Interaction between neocortical and hippocampal networks via slow oscillations. *Thalamus Relat. Syst.* **3**, 245–259 (2005).

63. Chung, J. E. et al. A fully automated approach to spike sorting. *Neuron* **95**, 1381–1394.e6 (2017).

64. Cowley, B. R. et al. DataHigh: graphical user interface for visualizing and interacting with high-dimensional neural activity. *J. Neural Eng.* **10**, 066012 (2013).

65. Schönemann, P. H. A generalized solution of the orthogonal procrustes problem. *Psychometrika* **31**, 1–10 (1966).

66. Gulati, T., Ramanathan, D. S., Wong, C. C. & Ganguly, K. Reactivation of emergent task-related ensembles during slow-wave sleep after neuroprosthetic learning. *Nat. Neurosci.* **17**, 1107–1113 (2014).

67. Eichenlaub, J. B. et al. Replay of learned neural firing sequences during rest in human motor cortex. *Cell Rep.* **31**, 107581 (2020).

68. Killick, R., Fearnhead, P. & Eckley, I. A. Optimal detection of changepoints with a linear computational cost. *J. Am. Stat. Assoc.* **107**, 1590–1598 (2012).

Acknowledgements

Research was supported by awards from the Department of Veterans Affairs, Veterans Health Administration (VA Merit: grant no. 1I01RX001640 to K.G.); the Weill Neurohub, the National Institute of Neurological Disorders and Stroke (grant no. 5K02NS093014 to K.G.; grant no. K99NS119737 to J.K.); and the American Heart Association Postdoctoral Fellowship (grant no. 831442 to J.K.). We thank S. J. Won for his help on the histology experiment.

Author information

Authors and Affiliations

1. Neurology Service, San Francisco Veterans Affairs Medical Center, San Francisco, CA, USA

   Jaekyung Kim & Karunesh Ganguly

2. Department of Neurology, University of California, San Francisco, San Francisco, CA, USA

   Jaekyung Kim & Karunesh Ganguly

3. HHMI and Departments of Physiology and Psychiatry, University of California, San Francisco, San Francisco, CA, USA

   Abhilasha Joshi & Loren Frank

Contributions

J.K., L.F. and K.G. conceived and designed the experiments. J.K. conducted the experiments and analysed the data. A.J. contributed to algorithm development for ripple detection. All contributed to the interpretation of the results. J.K. and K.G. wrote and A.J. and L.F. edited the manuscript.

Corresponding author

Correspondence to Karunesh Ganguly.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Validations of sleep and oscillation detections.

a, Histology image of a coronal brain section showing the depth of a recording electrode (left) located at the middle of the microwire arrays in M1 (right). **b**, Average of SO-triggered LFP (top; 10-15 Hz), and raster plot of an example unit of two rats; locked to the SO up-states. Mean ± s.e.m. **c**, Examples of detections of NREMS using power spectral density in 0.1-4 Hz and gamma (30-60 Hz) bands independently using the LFP of M1 (black, n = 77 sessions in 6 rats), PFC (red, n = 77 sessions in 6 rats), and hippocampus (HPC, blue, n = 77 sessions in 6 rats). No significant difference in NREM sleep duration across three recording areas (one-way ANOVA, $F_{2,231} = 1.47$,

P = 0.23; post hoc LME with two-sided *t*-test, corrected for multiple comparison, M1 vs. PFC: P = 0.13, M1 vs. HPC: P = 0.71, PFC vs. HPC: P = 0.24). Mean ± s.e.m. **d**, Average time course of ΔSO-SWR coupling (see Fig. 2) over the motor learning (n = 6 animals) using all detected NREM sleep (blue) versus using the first two tertiles (around early 66%) of the detected NREM sleep (yellow) in each sleep session. Mean ± s.e.m. **e**, Comparisons of rates between early period (n = 24 sessions in 6 rats) and late period (n = 24 sessions in 6 rats) in each corresponding metric (early days 1–4 vs. late days 10–13, LME with two-sided *t*-test; M1 SO: $t_{46} = -0.60$, P = 0.56; PFC SO: $t_{46} = 3.13$, P = 3.0×10^{-3}; SWRs: $t_{46} = 0.25$, P = 0.81). Mean ± s.e.m.

Extended Data Fig. 2 PFC-M1 SO coupling and changes in success rate.

a, Average time courses of PFC-M1 SO coupling during post-training sleep (bright red) and pre-training sleep (dark red). The time course in pre-training sleep was normalized referencing to the post-training sleep in each animal. Mean ± s.e.m. **b**, Comparison of single-day PFC-M1 SO coupling with the mean during late period of days 10–13 in the M1-SO-shuffled condition. Top horizontal lines: significant difference from the days 10–13; one-way ANOVA, $F_{12,218} = 1.73$, P = 0.063; post hoc LME with two-sided *t*-test, corrected for multiple comparison, **P = 9.2×10^{-3}. Mean ± s.e.m. **c**, Linear slopes of PFC-M1 SO coupling in a moving window. Gray-black traces represent measures with window sizes from 3-day (black) to 7-day (lightest gray). Blue represents average trace across the window sizes. Horizontal lines: significant difference from the days 10–13 using the 3-day size, LME with two-sided *t*-test, corrected for multiple comparison, ***P < 6.8×10^{-9}. Mean ± s.e.m. **d**, Computation of the PFC-M1 SO coupling during post-training sleep after subsampling PFC SOs to make the number of SOs used in the computation match across sessions (Methods). No session showed a significant difference between the observed condition for all PFC SOs and the subsampled condition; two-sided paired *t*-test, P > 0.28 for all sessions. Mean ± s.e.m. **e**, Computation of the changes in success rate (ΔSuccess rate) from the two-day history; it is shown using the task performance of an example animal. **f**, Pellet retrieval success rates (black) and changes in

success rate (magenta) in average across animals (n = 6 rats). Black dashed curve indicates single-exponential function fit ($R^2 = 0.85$, $P = 3.9 \times 10^{-11}$). Magenta dashed lines indicate piecewise linear regression fits. Mean ± s.e.m.

Extended Data Fig. 3 Comparis×ons in ΔSO-SWR coupling.

a, Comparison of distributions of $\Delta T_{SO\text{-}SWR}$ in M1 during post-training sleep in early learning (days 1–4). The SO-SWR coupling (probability of SWRs −0.75 to +0.75 sec from an SO, black box) during post-training sleep was compared with the SWR-shuffled condition (two-sided paired *t*-test, $t_{23} = -0.34$, P = 0.017). Mean ± s.e.m. **b**, Comparison of M1 ΔSO-SWR coupling during early (days 1—4) versus late period (days 10–13) in an individual animal. Mean of absolute M1 ΔSO-SWR coupling was significantly lower during the late compared to the early period; LME with two-sided *t*-test, P = 0.014, 0.049, 0.037, 0.011, 4.1×10^{-3}, and 1.1×10^{-3} for Rats 1–6. Mean ± s.e.m. **c**, Comparison of single-day M1 ΔSO-SWR coupling with the mean during late period of days 10—13 in the SWR-shuffled condition (i.e., temporal coupling was measured using the intact SOs and temporally disrupted structure of SWRs); one-way ANOVA, $F_{12,218} = 3.10$, $P = 4.5 \times 10^{-4}$; post hoc LME with two-sided *t*-test, corrected for multiple comparison, *P = 0.016. Mean ± s.e.m. **d**, Average time course of ΔSO-SWR coupling between PFC and hippocampus over the motor learning (gray; n = 6 rats); black is reproduced for M1 in Fig. 2d. Thick dashed curves indicate sigmoid function fits. Horizontal lines indicate the mean level during the two days prior to day 1. Mean ± s.e.m. **e**, Stage change point estimation using ‘change-point analysis’ for the single animal example in Fig. 2e. Black bars represent the estimated time point of stage transition. Dashed lines represent mean at each stage.

Extended Data Fig. 4 No change in cross-area coupling of spindles over learning.

a, Schematic model of the relationship across SWRs, SOs, and spindles. SOs

are coupled across PFC and M1 in relation to SWRs. These coordinated SOs also demonstrate increased communication across PFC-M1 (Fig. 4). Spindles are coupled to SOs in a local area processing but are not common during a cross-area coupled SO period. Our data suggest that the cross-area communication is related to SO-coupling and orchestration by SWR, while the spindle activity appears to be modulated by an area-specific process (e.g., corticothalamic connections projecting from each cortical area). **b**, Comparisons of the mean of triple phase-coupling (i.e., precise coupling of SWR, SO and spindle) in M1 (early days 1—4, n = 24 sessions in 6 rats: 7.4 ± 0.59% vs. late days 10–13, n = 24 sessions in 6 rats: 7.0 ± 0.75%, LME with two-sided *t*-test, $t_{46} = –0.47$, P = 0.64) and PFC (early days 1—4, n = 24 sessions in 6 rats: 6.3 ± 0.56% vs. late days 10–13, n = 24 sessions in 6 rats: 7.3 ± 0.69%, LME with two-sided *t*-test, $t_{46} = 1.23$, P = 0.25). Mean ± s.e.m.

Extended Data Fig. 5 Changes in spindles coupling with SO over learning.

a, Examples of the filtered LFP trace in M1 for SOs (0.1-4 Hz; same trace as in Fig. 1b) and spindles (10-15 Hz) during NREM sleep. SO and spindle are marked by gray boxes. Horizontal dashed lines indicate the threshold detecting SOs up/down states, and spindles onsets (lower dashed line) or high spindle power (upper dashed line). **b**, Schematic of SO-Spindle coupling, measured as the temporal lags of spindles from a SO up-state, $\Delta T_{\text{SO-Spindle}}$. SO-Spindle coupling was measured by coupling probability focused on the spindles from –0.5 to +1.0 sec from the time of the SO up-state (Methods). The difference of coupling between post-training sleep and pre-training sleep was examined over learning in '**c**' (ΔSO-Spindle coupling). **c**, Average time course of ΔSO-Spindle coupling in PFC (gray) and M1 (black) over the motor learning (n = 6 rats). Thick dashed curves indicate sigmoid function fits. Mean ± s.e.m. **d**, Comparisons of the mean of ΔSO-Spindle coupling in M1 (early days 1—4, n = 24 sessions in 6 rats: 1.3 ± 0.82% vs. late days 10–13, n = 24 sessions in 6 rats: –1.2 ± 0.62%, LME with two-sided *t*-test, $t_{46} = –2.99$, $P = 4.4 \times 10^{-3}$) and PFC (early days 1—3, n = 24 sessions in 6 rats: 1.18 ± 0.79% vs. late days 11–13, n = 24 sessions in 6 rats: –1.59 ± 0.98%, LME with two-sided *t*-test, $t_{46} = –2.12$, P = 0.041). Mean ± s.e.m.

Extended Data Fig. 6 Examples of neural trajectory templates and metrics of GPFA.

A, Examples of M1 neural trajectories traveling on the five GPFA factors in an animal. Gray curves with gradient intensity: mean templates of each session noted on the color bar. Triangles: reach onset (magenta) and mean pellet touch (cyan outline with filled by gray colors) time. **b**, Examples of mean neural trajectories traveling on the shared activity space defined by the top 3 GPFA factors in M1. Red curves: optimal neural trajectory for a successful task. Magenta triangles: reach onsets of each session. **c**, Average trace of shared-over-total variance (SOT) in M1 using all GPFA factors in the observed condition (black; early days 1–4, n = 24 sessions in 6 rats: 62.5 ± 2.6% vs. late days 10–13, n = 24 sessions in 6 rats: 71.4 ± 1.9%, LME with two-sided *t*-test, $t_{46} = 4.03$, $P = 2.0 \times 10^{-4}$) and the spike-shuffled condition (gray; early days 1–4, n = 24 sessions in 6 rats: 38.8 ± 2.4% vs. late days 10–13, n = 24 sessions in 6 rats: 42.6 ± 2.5 %, LME with two-sided *t*-test, $t_{46} = 1.54$, P = 0.13). n.s.: non-significance. Mean ± s.e.m. **d**, Average trace across animals for the mean (black) and the standard-deviation (s.d.; gray) of neural trajectory correlations in M1 (n = 6 rats). Mean ± s.e.m. **e**, Linear slopes of mean neural trajectories in a moving window in M1 and PFC. Gray-black traces represent measures with window sizes from 3-day (black) to 7-day (lightest gray). Blue represents average trace across the window sizes. Horizontal lines: significant difference from the days 1—4 using the 5-day size, LME with two-sided *t*-test, corrected for multiple comparison, *P = 0.032, ***$P < 7.6 \times 10^{-4}$ for M1, *P = 0.011, **$P = 3.2 \times 10^{-3}$ for PFC; results were similar with the other window size. Mean ± s.e.m.

Extended Data Fig. 7 Cross-area communications during long-term motor learning.

a, Comparison of distributions of PFC SOs from M1 SOs (ΔT_{SO}) during the late period of days 10–13 between the SWR^+ (1-sec epochs following SWRs onsets) and SWR^- condition (randomly selected 1-sec NREM epochs). No significant difference was found in each 50 ms bin between two conditions

using two-sided two-sampled *t*-test (n = 24 sessions in 6 rats). Mean ± s.e.m. **b**, Average of SWR-onset-triggered PETH in PFC and M1 (early days 1–4: n = 24 sessions in 6 rats; late days 10—13: n = 24 sessions in 6 rats). Mean ± s.e.m. **c**, Average trace of the correlation between PFC and M1 cross-area activities (Cross-area R) over the motor learning in the SWR^+ condition (early days 1—4, n = 24 sessions in 6 rats: 0.72 ± 0.033 vs. late days 10–13, n = 24 sessions in 6 rats: 0.84 ± 0.029, LME with two-sided *t*-test, $t_{46} = 3.41$, $P = 1.3 \times 10^{-4}$). Mean ± s.e.m. **d**, Comparisons of the percentage of quadruple phase-coupling (i.e., precise coupling of SWR, PFC SO, M1 SO, and spindle). There was no significant difference between the early versus the late period in M1 (early days 1—4, n = 24 sessions in 6 rats: 3.32 ± 0.34% vs. late days 10—13, n = 24 sessions in 6 rats: 3.87 ± 0.40%, LME with two-sided *t*-test, $t_{46} = 1.22$, P = 0.23) and PFC (early days 1—4, n = 24 sessions in 6 rats: 3.55 ± 0.41% vs. late days 10—13, n = 24 sessions in 6 rats: 3.85 ± 0.55%, LME with two-sided *t*-test, $t_{46} = 0.45$, P = 0.66). Mean ± s.e.m.

Extended Data Fig. 8 Parameters and changes over motor learning in M1 neural trajectory reactivations.

a, Schematic for finding optimal window size and time lag for the neural trajectory reactivation after a SWR onset. See the details in Methods for determining optimal window size. **b—c**, the distributions of window sizes and time lags of the final neural trajectory reactivations for the early period (899 SWRs during days 1—4, n = 6 rats) and the late period (826 SWRs during days 10–13, n = 6 rats). **d**, Average trace of reactivation correlation (R) over long-term motor learning (n = 6 rats) in the SWR^+ condition (early days 1–4, n = 24 sessions in 6 rats: 0.36 ± 0.023 vs. late days 10–13, n = 24 sessions in 6 rats: 0.42 ± 0.015, LME with two-sided *t*-test, $t_{46} = 3.53$, $P = 9.5 \times 10^{-4}$) versus the spike shuffled condition (early days 1–4, n = 24 sessions in 6 rats: 0.014 ± 0.012 vs. late days 10–13, n = 24 sessions in 6 rats: 0.0064 ± 0.0072, LME with two-sided *t*-test, $t_{46} = -0.42$, P = 0.68). n.s.: non-significance. Mean ± s.e.m. **e**, Comparisons of the mean of reactivation Rs between the SWR^+ versus the SWR^- condition during the late period of days 10–13 (n = 24 sessions in 6 rats, SWR^+: 0.39 ± 0.017 vs. SWR^-: 0.42 ± 0.015, two-sided paired *t*-test, $t_{23} = -2.69$, P = 0.013). Mean ± s.e.m. **f**, Comparisons

of the mean of reactivation Rs between the SWR^+SO^+ versus the SWR^+SO^- during the late period of days 10–13 (n = 24 sessions in 6 rats, SWR^+SO^+: 0.39 ± 0.018 vs. SWR^+SO^-: 0.39 ± 0.020, two-sided paired *t*-test, $t_{23} = 0.025$, P = 0.98). n.s.: non-significance. Mean ± s.e.m. **g**, Comparisons of the mean of all reactivation Rs in SWR^+ between pre- versus post-training sleep during the early (days 1–4, left; n = 24 sessions in 6 rats, pre-training: 0.35 ± 0.021 vs. post-training: 0.38 ± 0.020, two-sided paired *t*-test, $t_{23} = –2.14$, P = 0.043) and late period (days 10–13, right; n = 24 sessions in 6 rats, pre-training: 0.41 ± 0.014 vs. post-training: 0.42 ± 0.015, two-sided paired *t*-test, $t_{23} = –1.22$, P = 0.23). Mean ± s.e.m. **h**, Comparisons of the mean of top 10% reactivation Rs in SWR^+ between pre- versus post-training sleep during the early (days 1—4, left; n = 24 sessions in 6 rats, pre-training: 0.73 ± 0.020 vs. post-training: 0.77 ± 0.019, two-sided paired *t*-test, $t_{23} = –4.16$, $P = 4.1 \times 10^{-4}$) and late period (days 10–13, right; n = 24 sessions in 6 rats, pre-training: 0.81 ± 0.016 vs. post-training: 0.82 ± 0.014, two-sided paired *t*-test, $t_{23} = –1.48$, P = 0.15). Mean ± s.e.m.

Supplementary information

Reporting Summary

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Kim, J., Joshi, A., Frank, L. *et al.* Cortical–hippocampal coupling during manifold exploration in motor cortex. *Nature* **613**, 103–110 (2023). https://doi.org/10.1038/s41586-022-05533-z

- Received: 19 January 2022
- Accepted: 04 November 2022
- Published: 14 December 2022
- Issue Date: 05 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05533-z

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

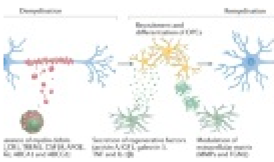
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

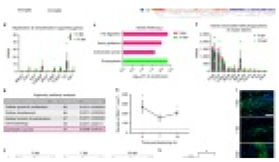
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05533-z



The pro-remyelination properties of microglia in the central nervous system

29 June 2019

Amy F. Lloyd & Veronique E. Miron



Central nervous system regeneration is driven by microglia necroptosis and repopulation

10 June 2019

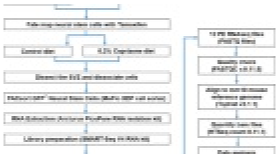
Amy F. Lloyd, Claire L. Davies, … Veronique E. Miron



Grey matter OPCs are less mature and less sensitive to IFNγ than white matter OPCs: consequences for remyelination

01 February 2018

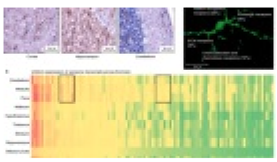
Dennis H. Lentferink, Jacomien M. Jongsma, … Wia Baron



[Transcriptomic analysis of loss of Gli1 in neural stem cells responding to demyelination in the mouse brain]

28 October 2021

Jayshree Samanta, Hernandez Moura Silva, … James L. Salzer



[Microglia in neurodegeneration]

26 September 2018

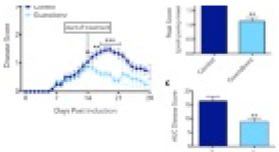
Suzanne Hickman, Saef Izzy, … Joseph El Khoury



[Fatal demyelinating disease is induced by monocyte-derived macrophages in the absence of TGF-β signaling]

16 April 2018

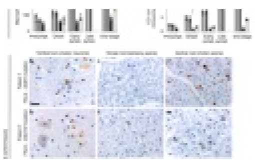
Harald Lund, Melanie Pieber, … Robert A. Harris



[Guanabenz modulates microglia and macrophages during demyelination]

09 November 2020

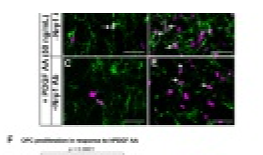
Kaitlyn Koenig Thompson & Stella E. Tsirka



[Modifying macrophages at the periphery has the capacity to change microglial reactivity and to extend ALS survival]

19 October 2020

Aude Chiot, Sakina Zaïdi, … Séverine Boillée



[Microglial neuropilin-1 promotes oligodendrocyte expansion during development and remyelination by trans-activating platelet-derived growth factor receptor]

15 April 2021

Amin Sherafat, Friederike Pfeiffer, … Akiko Nishiyama

- Article
- Open Access
- Published: 14 December 2022

Microglia regulate central nervous system myelin growth and integrity

- Niamh B. McNamara,
- David A. D. Munro,
- Nadine Bestard-Cuche,
- Akiko Uyeda,
- Jeroen F. J. Bogie,
- Alana Hoffmann,
- Rebecca K. Holloway,
- Irene Molina-Gonzalez,
- Katharine E. Askew,
- Stephen Mitchell,
- William Mungall,
- Michael Dodds,
- Carsten Dittmayer,
- Jonathan Moss,
- Jamie Rose,
- Stefan Szymkowiak,
- Lukas Amann,
- Barry W. McColl,
- Marco Prinz,
- Tara L. Spires-Jones,
- Werner Stenzel,
- Karen Horsburgh,
- Jerome J. A. Hendriks,
- Clare Pridans,
- Rieko Muramatsu,
- Anna Williams,

- Josef Priller &
- …
- Veronique E. Miron

Show authors

Nature **volume 613**, pages 120–129 (2023)

- 19k Accesses
- 221 Altmetric
- Metrics details

Abstract

Myelin is required for the function of neuronal axons in the central nervous system, but the mechanisms that support myelin health are unclear. Although macrophages in the central nervous system have been implicated in myelin health[1], it is unknown which macrophage populations are involved and which aspects they influence. Here we show that resident microglia are crucial for the maintenance of myelin health in adulthood in both mice and humans. We demonstrate that microglia are dispensable for developmental myelin ensheathment. However, they are required for subsequent regulation of myelin growth and associated cognitive function, and for preservation of myelin integrity by preventing its degeneration. We show that loss of myelin health due to the absence of microglia is associated with the appearance of a myelinating oligodendrocyte state with altered lipid metabolism. Moreover, this mechanism is regulated through disruption of the TGFβ1–TGFβR1 axis. Our findings highlight microglia as promising therapeutic targets for conditions in which myelin growth and integrity are dysregulated, such as in ageing and neurodegenerative disease[2,3].

Main

Myelin ensheathes neuronal axons to ensure their health and rapid

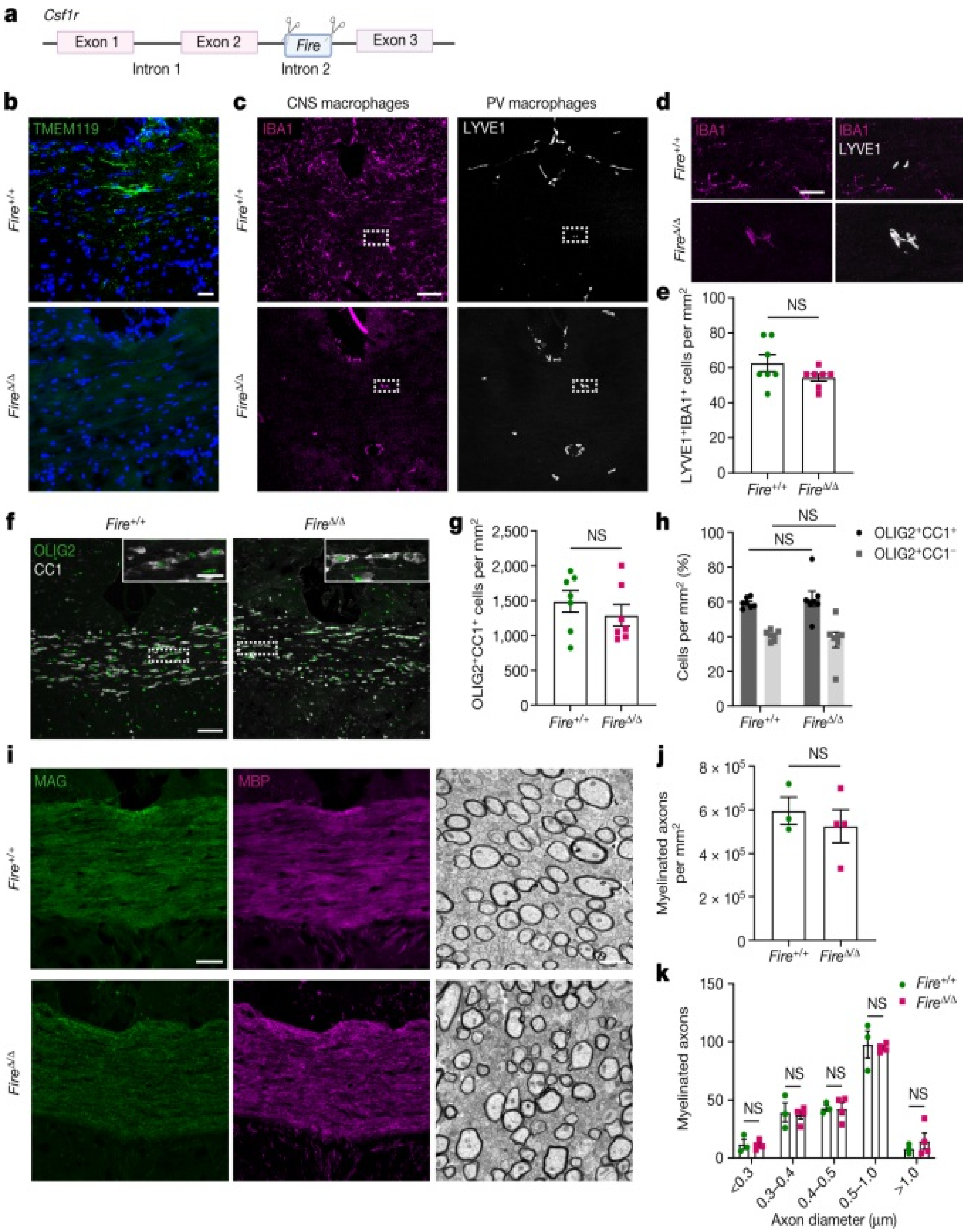
propagation of electrical impulses to support central nervous system (CNS) functions, for example, cognition. Learning and memory involve the formation of myelin and require myelin to be of good structural integrity. Myelin layers are compacted to a thickness proportional to the axon diameter[4]; however, with ageing and in neurodegenerative disease, disruption of these myelin properties occurs through hypermyelination. Enlarged areas of uncompacted myelin (where myelin grows) leads to myelin that is thicker, unravelling and forming protrusions (termed outfoldings), and loss of myelin integrity through degeneration also occurs in these contexts[2,3,5]. These myelin changes lead to impaired cognition in mice and predict poor cognitive performance in aged nonhuman primates and humans[6,7,8,9]. However, the fundamental mechanisms that coordinate appropriate formation, growth and integrity of CNS myelin are unclear. Recent research has implicated a population of CNS-resident macrophages, microglia, in this process. Myelination and the generation of myelin-forming oligodendrocytes are impaired following microglial depletion through loss of function of the pro-survival colony stimulating factor 1 receptor (CSF1R)[1]. However, this approach also targets CNS-resident border-associated macrophages (including perivascular macrophages) and blood monocytes. Therefore, it is unclear which macrophage populations regulate myelin, and the specific involvement of microglia in myelin formation and health is unknown.

Microglia are dispensable for myelination

To address these questions, we utilized a recently developed transgenic mouse model in which the Fms intronic regulatory element (*Fire*) super-enhancer of the *Csf1r* gene (*Fire*$^{\Delta/\Delta}$; Fig. 1a) is deleted. This deletion leads to an absence of microglia from development (when they normally emerge) through to adulthood, whereas other CNS macrophages are present[10,11]. These mice do not have many of the confounding issues that occur in other microglia-deficient models, such as developmental death, bone abnormalities and absence of CNS perivascular macrophages and monocytes[10]. Using the microglia marker TMEM119, we confirmed that microglia were depleted in the largest white matter tract of the brain, the corpus callosum (Fig. 1b). The few IBA1^{+} macrophages retained in *Fire*$^{\Delta/\Delta}$ mice were positive for the perivascular macrophage marker LYVE1 (Fig. 1c,d). Moreover, the presence

of these cells was confirmed by the observation of $CD206^{+}$ cells adjacent to $CD31^{+}$ blood vessels (Extended Data Fig. 1a). Perivascular macrophage densities were not significantly altered in *Fire*$^{\Delta/\Delta}$ mice at this time point (Fig. 1e) or at older ages (Extended Data Fig. 1b,c) despite the reduced expression of CSF1R (Extended Data Fig. 1d–f). Similar astrocyte numbers ($GFAP^{+}SOX9^{+}$) were observed in the corpus callosum of *Fire*$^{\Delta/\Delta}$ mice and *Fire*$^{+/+}$ littermates ($GFAP^{+}SOX9^{+}$ cells; Extended Data Fig. 1g,h). Notably, at an age when myelination is underway (postnatal day 25 (P25) to P30), *Fire*$^{\Delta/\Delta}$ mice had generated mature oligodendrocytes ($OLIG2^{+}CC1^{+}$) (Fig. 1f,g), which formed a similar proportion of the oligodendrocyte lineage ($OLIG2^{+}$) as in *Fire*$^{+/+}$ littermates (Fig. 1h). Myelin was formed in *Fire*$^{\Delta/\Delta}$ mice, as indicated by the expression of the myelin proteins MAG, MBP, CNPase, MOG and PLP in the corpus callosum (Fig. 1i and Extended Data Fig. 2a–c) and the cerebellum (Extended Data Fig. 2d,e). Ultrastructural analysis confirmed that myelination proceeded in *Fire*$^{\Delta/\Delta}$ mice (Fig. 1i), with no significant difference in the number of myelinated axons compared with *Fire*$^{+/+}$ mice (Fig. 1j) regardless of axon diameter (Fig. 1k). Our results indicate that microglia are dispensable for oligodendrocyte maturation and developmental myelin ensheathment. This finding is in contrast to previous attributions of these functions to microglia following depletion of all CNS macrophage populations.

Fig. 1: Microglia are not required for oligodendrocyte maturation and myelination.



a, *Fire*$^{\Delta/\Delta}$ mice were generated using CRISPR9–Cas9 deletion of the *Fire* super-enhancer located in intron 2 of *Csf1r*. **b**, Images of microglia

(TMEM119^{+}; green) in corpus callosum samples from *Fire*$^{+/+}$ and *Fire*$^{Δ/Δ}$ mice at 1 month of age, counterstained with Hoechst (blue). **c**, Images of CNS macrophages (IBA1^{+}; magenta) and perivascular (PV) macrophages (LYVE1^{+}; white) in *Fire*$^{+/+}$ and *Fire*$^{Δ/Δ}$ mice. **d**, Magnified images from **c** of IBA1^{+}LYVE1^{+} PV macrophages in *Fire*$^{+/+}$ and *Fire*$^{Δ/Δ}$ mice. **e**, Mean LYVE1^{+}IBA1^{+} cells per mm^{2} ± s.e.m. in *Fire*$^{+/+}$ and *Fire*$^{Δ/Δ}$ mice. $n = 7$ mice per group. $P = 0.1411$, two-tailed unpaired Student's *t*-test. **f**, Images of mature oligodendrocytes expressing both OLIG2 (green) and CC1 (white) in *Fire*$^{+/+}$ and *Fire*$^{Δ/Δ}$ mice. Inset shows magnified view. **g**, Mean OLIG2^{+}CC1^{+} cells per mm^{2} ± s.e.m. in *Fire*$^{+/+}$ and *Fire*$^{Δ/Δ}$ mice. $n = 7$ mice per group. $P = 0.1990$, two-tailed unpaired Student's *t*-test. **h**, Mean proportion of cells of the oligodendrocyte lineage (OLIG2^{+}), which are mature (CC1^{+}; black) or immature (CC1^{-}; grey) (±s.e.m.). $n = 7$ mice per group. CC1^{+}, $P = 0.9472$; CC1^{-}, $P = 0.9472$; one-way analysis of variance (ANOVA) with Tukey's multiple comparisons test. **i**, Images of corpus callosum from *Fire*$^{+/+}$ and *Fire*$^{Δ/Δ}$ mice stained for the myelin proteins MAG (green) and MBP (magenta) (left and middle) and imaged by electron microscopy (right). **j**, Mean number of myelinated axons per mm^{2} ± s.e.m. in corpus callosum from *Fire*$^{+/+}$ and *Fire*$^{Δ/Δ}$ mice. $n = 3$ *Fire*$^{+/+}$ mice and 4 *Fire*$^{Δ/Δ}$ mice. $P = 0.5216$, two-tailed unpaired Student's *t*-test. **k**, Mean number of myelinated axons in corpus callosum from *Fire*$^{+/+}$ and *Fire*$^{Δ/Δ}$ mice per axon diameter. $n = 3$ *Fire*$^{+/+}$ mice and 4 *Fire*$^{Δ/Δ}$ mice. $P = 0.9139$, two-way ANOVA with Sidak's multiple comparisons test. Scale bars, 25 µm (**b**,**d**,**i** (left and middle)), 75 µm (**c**,**f**) and 1 µm (**i** (right)).

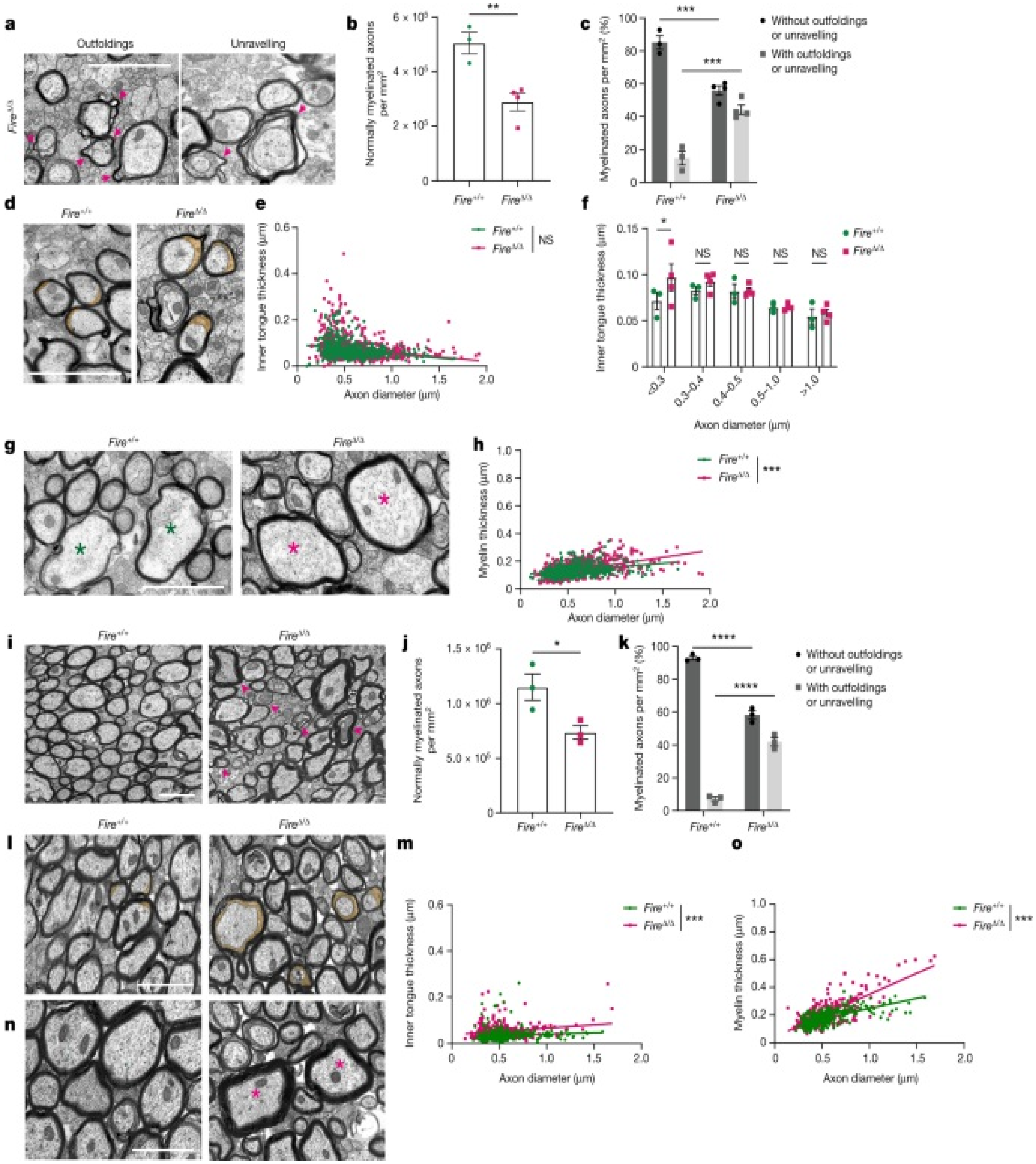
Source data

Microglia prevent hypermyelination

However, *Fire*$^{Δ/Δ}$ mice showed abnormal myelin structure indicative of hypermyelination. At P25–P30, there was an increase in myelin that was outfolding or unravelling in *Fire*$^{Δ/Δ}$ mice (Fig. 2a–c), which was documented in 44% of sheaths compared with only 14% in *Fire*$^{+/+}$ controls. *Fire*$^{Δ/Δ}$ mice showed enlarged areas of uncompacted myelin (inner tongue) on smaller diameter axons (Fig. 2d–f). As enlarged inner tongues precluded

conventional *g* ratio analysis, we measured myelin thickness directly and observed increased myelin thickness in *Fire*$^{\Delta/\Delta}$ mice preferentially on large diameter axons (Fig. 2g,h). Axon diameter-dependent observations may reflect that myelination of larger diameter axons occurs before that of smaller ones, first involving growth at the inner tongue followed by compaction that thickens the myelin sheath. We assessed the impact of these myelin changes on axonal health in *Fire*$^{\Delta/\Delta}$ mice. Although expression of phosphorylated neurofilament was unaffected at 1 month of age, axonal spheroids, which are indicative of impaired axonal transport, were occasionally observed at later ages in <0.1% of myelinated axons (Extended Data Fig. 2f–h). Myelin outfoldings and unravelling persisted in *Fire*$^{\Delta/\Delta}$ mice at 3–4 months of age (Fig. 2i–k), and enlarged inner tongues and increased myelin thickness were observed across all axon diameters compared with *Fire*$^{+/+}$ mice (Fig. 2l–o) and younger *Fire*$^{\Delta/\Delta}$ mice (Extended Data Fig. 3). Therefore, hypermyelination occurs in the absence of microglia, which indicates that microglia are required for the regulation of myelin growth.

Fig. 2: Microglia regulate myelin growth.



Assessments at 1 month (**a**–**h**) and 3–4 months (**i**–**o**) of age in *Fire*$^{+/+}$ and *Fire*$^{\Delta/\Delta}$ mice. **a**, Images of myelin abnormalities (arrowheads) in *Fire*$^{\Delta/\Delta}$ mice. **b**, Mean number of normally myelinated axons per mm^2 ± s.e.m. $n = 3$ *Fire*$^{+/+}$ mice and 4 *Fire*$^{\Delta/\Delta}$ mice. $^{**}P = 0.008$, two-tailed unpaired Student's *t*-test. **c**, Mean proportion of axons with and without outfoldings and

unravelling per $mm^2 \pm$ s.e.m. $n = 3$ $Fire^{+/+}$ mice and 4 $Fire^{\Delta/\Delta}$ mice. With, ***$P = 0.0006$; without, ***$P = 0.0006$; one-way ANOVA with Tukey's multiple comparisons test. **d**, Images of inner tongues (orange). **e**, Inner tongue thickness (μm) versus axon diameter. $n = 200$ axons per mouse, 3 $Fire^{+/+}$ mice and 4 $Fire^{\Delta/\Delta}$ mice. **f**, Mean inner tongue thickness per axon diameter. $n = 3$ $Fire^{+/+}$ mice and 4 $Fire^{\Delta/\Delta}$ mice. <0.3 μm, *$P = 0.0417$, two-way ANOVA with Sidak's multiple comparisons test. **g**, Images of myelin thickness (asterisks indicate axons of similar size). **h**, Myelin thickness versus axon diameter. $n = 200$ axons per mouse, 3 $Fire^{+/+}$ mice and 4 $Fire^{\Delta/\Delta}$ mice. ***$P < 0.0001$, simple linear regression of slopes. **i**, Images of myelin abnormalities (arrowheads) in $Fire^{\Delta/\Delta}$ mice. **j**, Mean number of normally myelinated axons per $mm^2 \pm$ s.e.m. $n = 3$ mice per group. *$P = 0.0372$, two-tailed unpaired Student's *t*-test. **k**, Mean proportion of axons with and without outfoldings and unravelling per $mm^2 \pm$ s.e.m. $n = 3$ mice per group. With, ****$P < 0.0001$; without, ****$P < 0.0001$; one-way ANOVA with Tukey's multiple comparisons test. **l**, Images of inner tongues (orange). **m**, Inner tongue thickness versus axon diameter. $n = 100$ axons per mouse, 3 mice per group. ***$P < 0.0001$, simple linear regression of intercepts. **n**, Images of myelin thickness (asterisks indicate axons of similar size). **o**, Myelin thickness versus axon diameter. $n = 100$ axons per mouse, 3 mice per group. ***$P < 0.0001$, simple linear regression of slopes. Scale bars, 1 μm (**a**,**d**,**g**,**i**,**l**,**n**).

Source data

Given that these changes in myelin structure are sufficient to cause cognitive impairment in other models[12,13], we evaluated cognition in $Fire^{\Delta/\Delta}$ mice using the Barnes maze spatial learning and memory task (Extended Data Fig. 4a). Both $Fire^{\Delta/\Delta}$ and $Fire^{+/+}$mice became progressively faster at locating the target hole with the underlying escape chamber (primary latency), which indicated spatial learning in these mice (Extended Data Fig. 4b,c). Following removal of the escape chamber, probes 1 h and 3 days later indicated no memory-encoding deficit, as indicated by the percentage of time spent in the target quadrant and the number of nose pokes in and around the target hole (Extended Data Fig. 4d–f). Next, we tested cognitive flexibility, which is learning to adjust thinking from an old to a new situation and is highly

dependent on the structural integrity of myelin[6,14,15]. This experiment involves learning to locate an escape hole placed 180° from the original target (Extended Data Fig. 4g–k). Although the time taken to locate the new target was unimpaired in *Fire*$^{\Delta/\Delta}$ mice, significantly more errors were made before reaching it (Extended Data Fig. 4h,i), which indicated that these mice have poor cognitive flexibility. *Fire*$^{\Delta/\Delta}$ mice did not have confounding anxiety or motor deficits (Extended Data Fig. 4l–r). Therefore, the absence of microglia is associated with impaired cognitive flexibility.

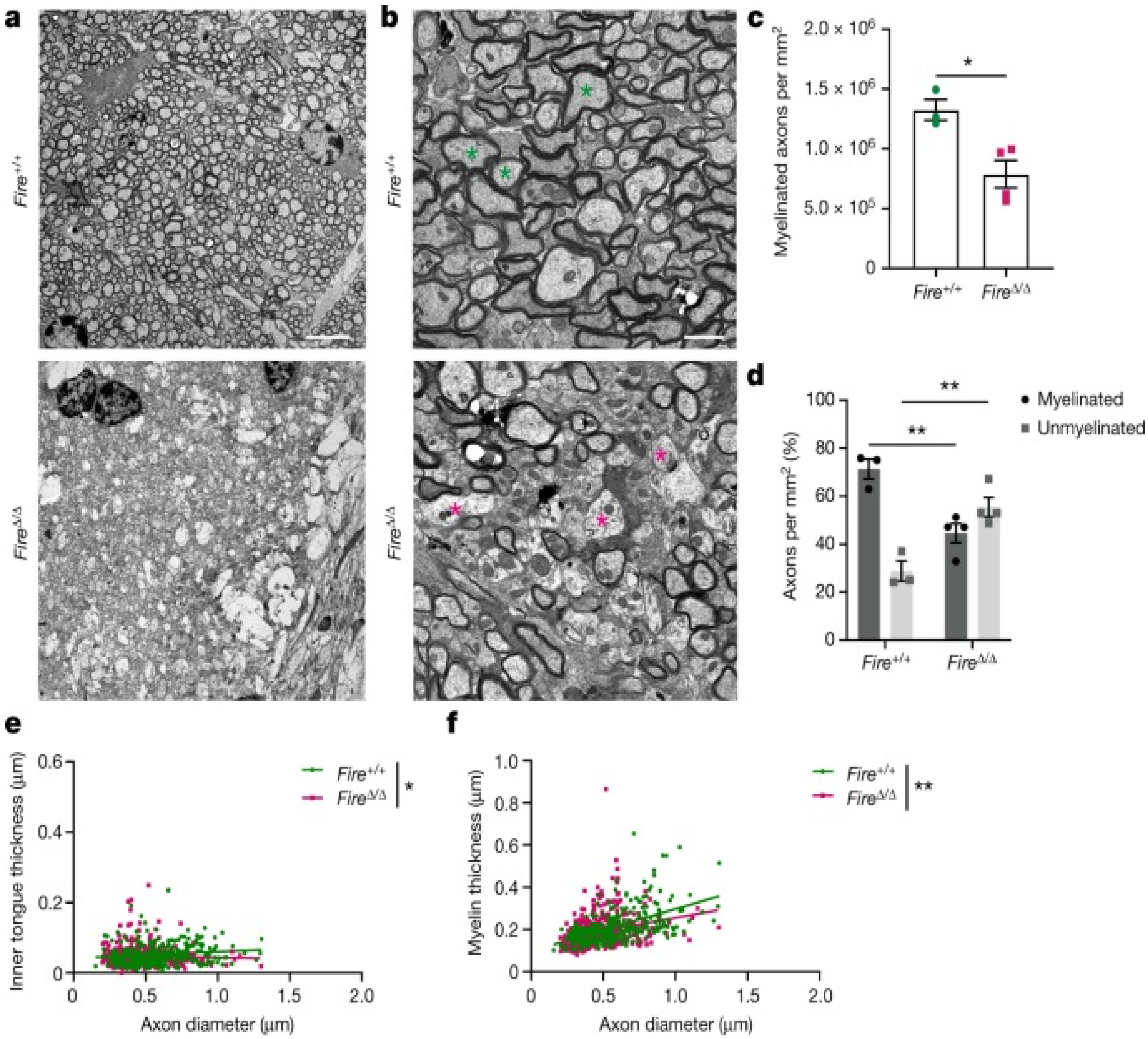
Recent studies have indicated that learning and memory encoding require new oligodendrogenesis, and that long-term consolidation of this information involves increased myelination[16,17]. Therefore, we assessed whether these processes occur in the absence of microglia. The generation of new oligodendrocytes from proliferating progenitor cells was identified through the incorporation of 5-ethynyl-2′-deoxyuridine (EdU) provided during the cognitive testing stage. There was a similar number of newly generated oligodendrocytes in *Fire*$^{\Delta/\Delta}$ mice and *Fire*$^{+/+}$ mice (Extended Data Fig. 5a–d), which was consistent with the largely unimpaired learning and memory encoding observed in *Fire*$^{\Delta/\Delta}$ mice. However, whereas *Fire*$^{+/+}$ mice had a significantly increased number of myelinated axons in the corpus callosum 6 weeks after completion of the cognitive task, this did not differ between untrained and trained *Fire*$^{\Delta/\Delta}$ mice (Extended Data Fig. 5e–g). These findings suggest that the absence of microglia prevents the increase in myelination that normally occurs with consolidation of new spatial information. Of note, we did not find an association between the number of myelinated axons in a given mouse and its reversal cognitive performance (Extended Data Fig. 5h). This result suggests that either a threshold number of myelinated axons is required for cognitive flexibility or that myelin structural changes may be more relevant for this function.

Microglia prevent demyelination

Assessment of myelin in *Fire*$^{\Delta/\Delta}$ mice at 6 months of age showed areas of substantial demyelination (Fig. 3a) and areas of patchy demyelination. This in turn resulted in a significant decrease in the number and proportion of myelinated axons compared with *Fire*$^{+/+}$ mice (Fig. 3b–d). The patchy nature

of demyelination did not result in widespread loss of myelin protein across the corpus callosum (Extended Data Fig. 6a–c). Axonal spheroids were rarely observed (<0.1% of axons). Axons retaining myelin in 6-month-old *Fire*$^{\Delta/\Delta}$ mice had reduced inner tongue size and thinner myelin compared with *Fire*$^{+/+}$ mice (Fig. 3e,f and Extended Data Fig. 6d) and with 3–4-month-old *Fire*$^{\Delta/\Delta}$ mice (Extended Data Fig. 3). Demyelination was not associated with loss of oligodendrocytes at this age or younger (Extended Data Fig. 6e–h). Demyelination was initiated at 4.5 months of age in *Fire*$^{\Delta/\Delta}$ mice (Extended Data Fig. 6i–l), and unmyelinated axons were medium-to-large calibre (mean 0.73 µm ± 0.1 s.e.m.) in size (Extended Data Fig. 6l). This result indicated that these axons underwent demyelination first, as medium-to-large diameter axons showed hypermyelination immediately before demyelination at 3–4 months of age (Extended Data Fig. 6k); therefore, hypermyelination may precede demyelination. These findings demonstrate that the lack of microglia is sufficient to induce CNS demyelination with increasing age.

Fig. 3: Absence of microglia causes demyelination.



a, Images of substantial demyelination in the corpus callosum of 6-month-old *Fire*$^{\Delta/\Delta}$ mice compared with age-matched *Fire*$^{+/+}$ mice. Scale bar, 5 µm. **b**, Images of patchy demyelination in *Fire*$^{\Delta/\Delta}$ mice compared with age-matched *Fire*$^{+/+}$ mice. Asterisks indicate axons of similar size. Scale bar, 1 µm. **c**, Mean number of myelinated axons per mm^2 ± s.e.m. $n = 3$ *Fire*$^{+/+}$ mice and 4 *Fire*$^{\Delta/\Delta}$ mice. *$P = 0.0170$, two-tailed unpaired Student's *t*-test. **d**, Mean proportion of axons that are myelinated (black) and unmyelinated (grey) per mm^2 ± s.e.m. $n = 3$ *Fire*$^{+/+}$ mice and 4 *Fire*$^{\Delta/\Delta}$ mice. Myelinated, **$P = 0.0051$; unmyelinated, **$P = 0.0051$; one-way ANOVA with Tukey's multiple comparisons test. **e**, Inner tongue thickness versus axon diameter. $n = 100$ axons per mouse, 3 *Fire*$^{+/+}$ mice and 4 *Fire*$^{\Delta/\Delta}$ mice. *$P = 0.0332$,

simple linear regression of slopes. **f**, Myelin thickness versus axon diameter. $n = 100$ axons per mouse, 3 *Fire*$^{+/+}$ mice and 4 *Fire*$^{\Delta/\Delta}$ mice. $^{**}P = 0.0062$, simple linear regression of slopes.

Source data

Microglia maintain existing myelin

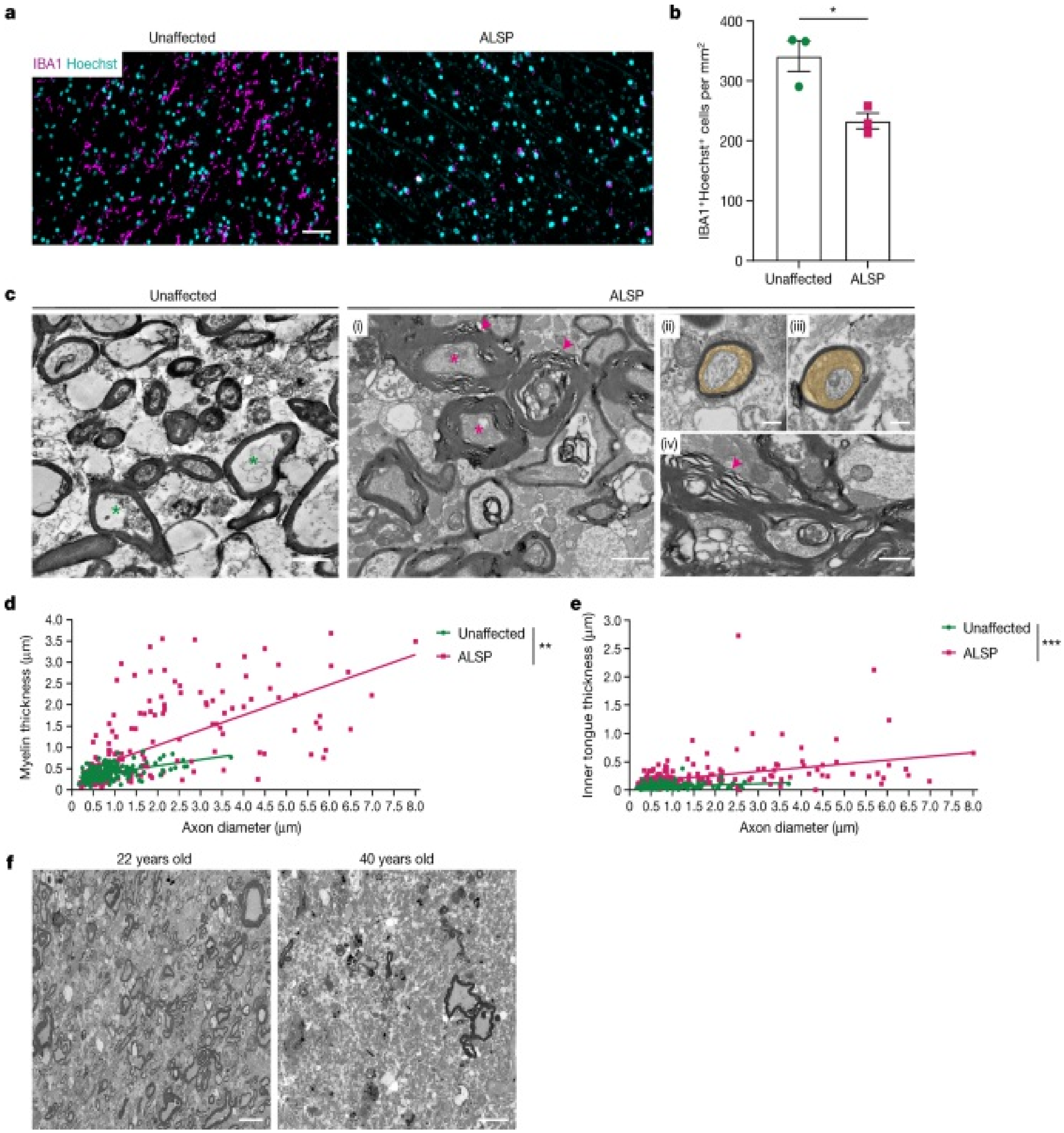
We next asked whether these changes in myelin growth and integrity reflect disruption of myelin formation or myelin maintenance. To that end, we depleted microglia in mice when developmental myelination is complete (2 months of age onwards) by providing the CSF1R inhibitor PLX5622 in the diet of adult *Fire*$^{+/+}$ mice for 1 month. This resulted in >50% reduction of IBA1^{+} cells at 3 months of age (Extended Data Fig. 7a–c). Compared with mice fed the control diet, microglia depletion from 2 to 3 months of age resulted in enlarged inner tongues and thicker myelin (Extended Data Fig. 7d–g), whereas depletion from 5 to 6 months of age caused patchy demyelination (Extended Data Fig. 7j–l). Oligodendrocyte numbers were unchanged (Extended Data Fig. 7h,i). Therefore, microglia depletion in adult mice mirrored the hypermyelination and myelin degeneration observed at equivalent ages in *Fire*$^{\Delta/\Delta}$ mice, which indicated that microglia are required for myelin maintenance once it is already formed.

Microglia deficits in humans affect myelin health

After demonstrating that microglia are required for myelin health in mice, we investigated the relevance of these findings in humans. We analysed samples from individuals with the rare leukoencephalopathy adult-onset leukoencephalopathy with axonal spheroids and pigmented glia (ALSP) (Extended Data Table 1). In ALSP, heterozygous *CSF1R* mutations lead to cognitive dysfunction in association with reduced IBA1^{+} parenchymal cells, especially in frontal white matter, whereas those in the grey matter are relatively preserved[18]. In comparison to age-matched individuals who died of non-neurological causes, there was a significant decrease in IBA1^{+} microglia and macrophages in the frontal white matter of individuals with ALSP (Fig. 4a,b). Moreover, there was a relative increase in the proportion of

perivascular macrophages ($IBA1^{+}LYVE1^{+}$) (Extended Data Fig. 8a–d). Ultrastructural analysis of ALSP white matter revealed myelin outfoldings and unravelling (Fig. 4c), thicker myelin (Fig. 4c,d) and enlarged inner tongues (Fig. 4c,e and Extended Data Fig. 8e–g). Demyelination was also observed and progressively worsened with age (Fig. 4f). Larger axon diameters were noted in ALSP samples compared with unaffected samples (Fig. 4d,e). This result is consistent with axonal swelling being a typical pathological feature of this disorder; however, myelin was still thicker than would be expected of these axon diameters. Extra thick myelin in ALSP was associated with myelin unravelling (Fig. 4c), which may indicate early stages of a transition from hypermyelination to demyelination. These findings show that a reduction in white matter microglia in humans is associated with hypermyelination and eventual demyelination.

Fig. 4: Reduction of microglia in human white matter is associated with hypermyelination and demyelination.



a, Images of IBA1^{+} macrophages (magenta) in human frontal white matter from individuals with ALSP and unaffected age-matched individuals, counterstained with Hoechst (cyan). **b**, Mean IBA1^{+} cells per mm^{2} ± s.e.m. in samples from unaffected individuals and individuals with ALSP. $n = 3$ samples per group. $*P = 0.0197$, two-tailed unpaired Student's *t*-test. **c**, Images of frontal white matter in unaffected and ALSP samples (**i**). Asterisks indicate axons of similar size, arrowheads indicate myelin abnormalities.

Panels (ii) and (iii) show enlarged inner tongues in ALSP (orange). Panel (iv) shows myelin outfoldings and unravelling in ALSP (arrow). **d**, Myelin thickness versus axon diameter. $n = 100$ axons per sample, $n = 2$ samples per group. $^{**}P = 0.002$, simple linear regression of slopes. **e**, Inner tongue thickness (µm) versus axon diameter. $n = 100$ axons per sample, $n = 2$ samples per group. $^{***}P < 0.0001$, simple linear regression of intercepts. **f**, Images of extent of demyelination of frontal white matter in individuals with ALSP; 22-year-old individual and 40-year-old individual. Scale bars, 0.5 µm (**c** (ii)–(iv)), 1 µm (**c** (i)). 10 µm (**f**) or 50 µm (**a**).

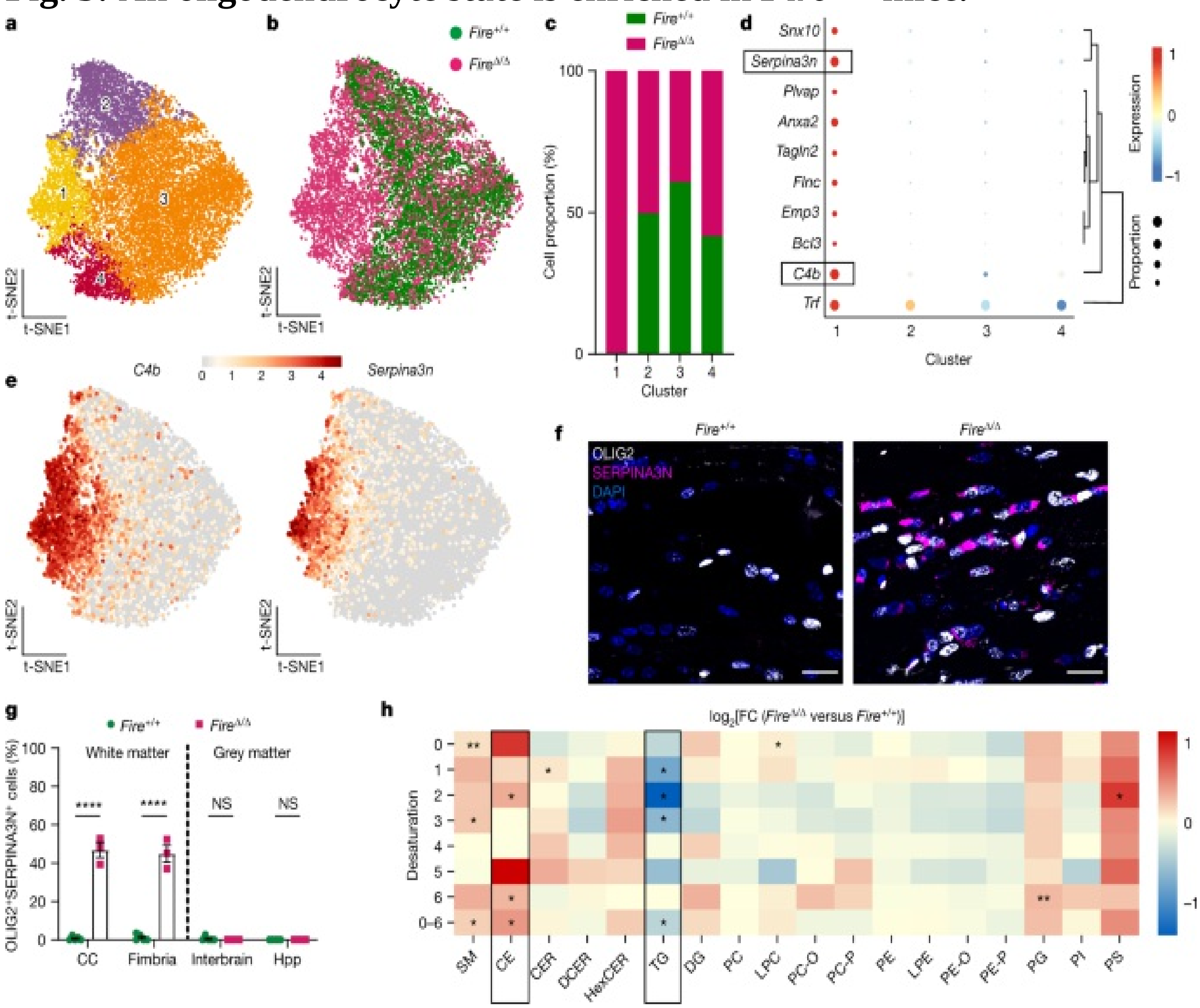
Source data

Microglia suppress oligodendrocyte state

We next sought to determine the cellular and molecular mechanisms that underpin the loss of myelin health in the absence of microglia. To that end, we performed single-cell RNA sequencing of brain samples from $Fire^{\Delta/\Delta}$ and $Fire^{+/+}$ mice at 1 month of age (Extended Data Fig. 9a–j). Mature oligodendrocytes were identified through the expression of myelin genes (*Plp*, *Mog*, *Mag* and *Mbp*) and the absence of expression of markers for other cell types (Extended Data Fig. 9a–d). The oligodendrocytes were clustered into four states (Oligo1 to Oligo4) (Fig. 5a and Extended Data Fig. 9k,l). Notably, we observed that cluster 1 oligodendrocytes (Oligo1) were almost exclusively found in $Fire^{\Delta/\Delta}$ mice (Fig. 5b,c) and distinguished by the high expression of genes (Supplementary Table 1 and Fig. 5d) including *Serpina3n* and *C4b* (Fig. 5d,e). We confirmed the presence of $SERPINA3N^{+}OLIG2^{+}$ cells almost exclusively in $Fire^{\Delta/\Delta}$ mouse white matter, whereas these cells were undetectable in grey matter of either genotype (Fig. 5f,g). Analysis of differentially expressed genes in the Oligo1 cluster revealed that the top canonical pathways were related to lipid synthesis and metabolism. Specifically, ingenuity pathway analysis highlighted the following significant pathways: superpathway of cholesterol biosynthesis ($P = 8 \times 10^{-12}$) and cholesterol biosynthesis I–III ($P = 1.23 \times 10^{-7}$). Analysis using the DAVID bioinformatics resource highlighted the following pathways: lipid biosynthesis ($P = 3.8 \times 10^{-8}$); lipid metabolism ($P = 5.1–7.0 \times 10^{-6}$); and cholesterol metabolism ($P = 9.2 \times 10^{-4}$). Of note,

cholesterol is enriched in myelin and required for myelin growth[19]. Lipidomics analysis of *Fire*$^{\Delta/\Delta}$ mouse white matter revealed an increase in cholesterol esters and a decrease in triglycerides (Fig. 5h), which was indicative of excess cholesterol and impaired lipid export, respectively. This result is consistent with the observed surplus in myelin membrane formation in *Fire*$^{\Delta/\Delta}$ mice. Moreover, dysregulation of genes associated with cholesterol transport has been reported in ALSP[20].

Fig. 5: An oligodendrocyte state is enriched in *Fire*$^{\Delta/\Delta}$ mice.



a, T-distributed stochastic neighbour embedding (t-SNE) plots of oligodendrocyte clusters in *Fire*$^{+/+}$ and *Fire*$^{\Delta/\Delta}$ mice: Oligo1 (yellow), Oligo2 (purple), Oligo3 (orange) and Oligo4 (red). **b**, Distribution of oligodendrocyte clusters between *Fire*$^{+/+}$ (green) and *Fire*$^{\Delta/\Delta}$ (magenta) mice.

c, Oligodendrocyte cluster proportions in *Fire*$^{+/+}$ and *Fire*$^{\Delta/\Delta}$ mice. $n = 4$ mice per group. **d**, Top differentially expressed genes in Oligo1 compared with Oligo2 to Oligo4. Level of normalized expression of each gene is indicated by the heatmap: 1, red, −1, blue. Proportion of cells expressing each gene is indicated by the size of the circles in the plot. **e**, t-SNE projection of expression of *C4b* and *Serpina3n*. **f**, Images of OLIG2^{+} cells (white) expressing SERPINA3N (magenta) and counterstained with DAPI (blue) in the corpus callosum. **g**, Mean percentage of OLIG2^{+} cells expressing SERPINA3N ± s.e.m. in white matter (corpus callosum (CC) and fimbria) and grey matter (interbrain and hippocampus (Hpp)). $n = 5$ *Fire*$^{+/+}$ mice and 3 *Fire*$^{\Delta/\Delta}$ mice. CC and fimbria, ****$P < 0.0001$; interbrain and Hpp, $P = 0.9989$ and $P > 0.9999$, respectively; two-way ANOVA with Sidak's multiple comparisons test. **h**, Lipidomics analysis represented as $\log_2$(fold change (FC)) in *Fire*$^{\Delta/\Delta}$ mice versus *Fire*$^{+/+}$ mice, with upregulated lipid species indicated in red and downregulated lipid species indicated in blue, ordered based on desaturation of fatty acids (double bonds) and total class value (0–6). Boxes indicate lipid species of interest. SM, sphingomyelins; CE, cholesterol esters; CER, ceramides; DCER, dihydroceramides; HexCER, hexosylceramides; TG, triaglycerides; DG, diacylglycerides; PC, phosphatidylcholine; LPC, lysophosphatidylcholine; PC-O, 1-alkyl,2-acylphosphatidylcholines; PC-P, 1-alkenyl,2-acylphosphatidylcholines; PE, phosphatidylethanolamine; LPE, lysophosphatidylethanolamine; PE-O, 1-alkyl,2-acylphosphatidylethanolamines; PE-P, 1-alkenyl,2-acylphosphatidylethanolamines; PG, phosphatidylglycerol; P I, phosphatidylinositol; PS, phosphatidylserine. $n = 3$ mice per group. One-sample *t*-test of $\log_2$(FC) against a value of 0: SM-0, **$P = 0.0058$; SM-3, *$P = 0.0202$; SM-0–6, *$P = 0.0347$; CE-2, *$P = 0.0384$; CE-6, *$P = 0.0474$; CE-0–6, *$P = 0.0276$; CER-1, *$P = 0.0171$; TG-1, *$P = 0.0301$; TG-2, *$P = 0.0116$; TG-3, *$P = 0.0432$; TG-0–6, *$P = 0.0412$; LPC-0, *$P = 0.0415$; PG-6, **$P = 0.0055$; PS-2, *$P = 0.0459$.

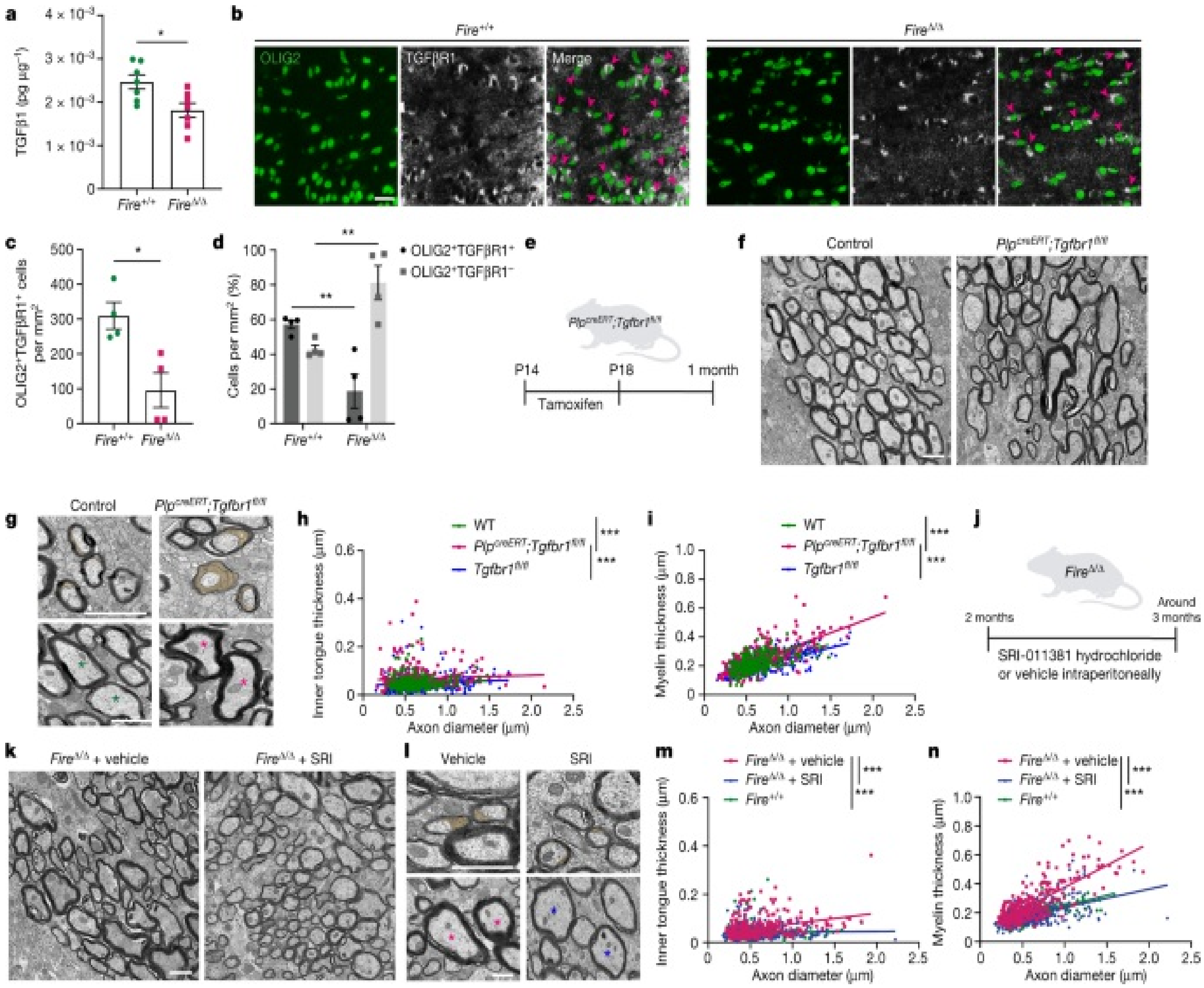
Source data

TGFβ1 signalling regulates myelin health

To determine how the absence of microglia contributes to these findings, we

assessed the predicted upstream regulators of genes in the Oligo1 cluster. TGFβ1 was identified as a prime candidate ($P = 7.7 \times 10^{-13}$; Supplementary Table 2), as it is predominantly expressed by microglia in both mouse and human brain (https://www.brainrnaseq.org)[21,22], and it is known to influence lipid metabolism[23]. Accordingly, TGFβ1 levels were significantly downregulated in *Fire*$^{\Delta/\Delta}$ mouse white matter (Fig. 6a). We next assessed the capacity of oligodendrocytes to respond to TGFβ1. Although there were abundant TGFβR1^{+}OLIG2^{+} cells in *Fire*$^{+/+}$ mice, the number and percentage of these cells were significantly reduced in *Fire*$^{\Delta/\Delta}$ mice (Fig. 6b–d). These findings led us to ask whether elimination of TGFβR1 signalling in oligodendrocytes is sufficient to cause myelin pathology. As *Tgfb1* knockout in the CNS results in a confounding decrease in microglia number, loss of microglia homeostasis and monocyte infiltration[24], we utilized a conditional knockout of *Tgfbr1* in mature oligodendrocytes (*Plp*^{creERT};*Tgfbr1*$^{fl/fl}$). Following tamoxifen administration from P14 to P18 (Fig. 6e), TGFβR1 expression by oligodendrocyte lineage cells was significantly reduced in *Plp*^{creERT};*Tgfbr1*$^{fl/fl}$ mice at 1 month of age compared with controls (Extended Data Fig. 10a,b). There was no significant impact on the number of myelinated axons (Extended Data Fig. 10c). However, by P28, conditional knockout mice had enlarged inner tongues on smaller diameter axons (Fig. 6f–h and Extended Data Fig. 10d) and thicker myelin on larger diameter axons (Fig. 6f,g,i and Extended Data Fig. 10e) relative to tamoxifen-treated floxed and wild-type controls. This result mimics the observations in *Fire*$^{\Delta/\Delta}$ mice at 1 month of age.

Fig. 6: TGFβR1 signalling regulates myelin integrity.



a, TGFβ1 levels (pg μg^{-1} ± s.e.m.) in corpus callosum normalized to the respective total protein. n = 7 mice per group. *P = 0.0150, two-tailed unpaired Student's t-test. **b**, Images of OLIG2^{+} cells (green) expressing TGFβR1 (white), (arrowheads) in 1-month-old mice. **c**, Mean OLIG2^{+}TGFβR1^{+} cells ± s.e.m. n = 4 mice per group. *P = 0.0144, two-tailed unpaired Student's t-test. **d**, Mean percentage of OLIG2^{+} cells that are TGFβR1^{+} (black) or TGFβR1^{-} (grey) (± s.e.m.), *P = 0.0120 and *P = 0.0120, respectively, n = 3 mice per group. One-way ANOVA with Tukey's multiple comparisons test. **e**, Mice were treated with tamoxifen from P14 to P18 then assessed at 1 month. **f**, Images of *Plp*^{creERT}*;Tgfbr1*$^{fl/fl}$ mice versus control. **g**, Images of *Plp*^{creERT}*;Tgfbr1*$^{fl/fl}$ mice, indicating enlarged inner tongues (orange) and hypermyelination (asterisks) compared with control. **h**, Inner tongue thickness versus axon diameter. n = 100 axons per mouse, 3

mice per group. ***$P < 0.0001$, simple linear regression of intercepts. *Plp*^{creERT};*Tgfbr1*$^{fl/fl}$ versus wild type and versus *Tgfbr1*$^{fl/fl}$, ***$P < 0.0001$; wild type versus *Tgfbr1*$^{fl/fl}$, $P = 0.3228$; Kruskal–Wallis with Dunn's multiple comparisons test. **i**, Myelin thickness versus axon diameter. $n = 100$ axons per mouse, 3 mice per group. ***$P < 0.0001$, simple linear regression of slopes. *Plp*^{creERT};*Tgfbr1*$^{fl/fl}$ versus wild type, ***$P = 0.0008$ and versus *Tgfbr1*$^{fl/fl}$, ***$P = 0.0001$; wild type versus *Tgfbr1*$^{fl/fl}$, $P > 0.9999$ Kruskal–Wallis with Dunn's multiple comparisons test. **j**, *Fire*$^{\Delta/\Delta}$ mice were treated with SRI-011381 hydrochloride (SRI; 30 mg kg^{-1}) or vehicle from 2–3 months of age. **k**, Images of *Fire*$^{\Delta/\Delta}$ mice treated with vehicle or SRI-011381. **l**, Images of *Fire*$^{\Delta/\Delta}$ mice treated with vehicle or SRI-011381 indicating inner tongue size (orange) and myelin thickness (asterisks). **m**, Inner tongue thickness versus axon diameter. $n = 100$ axons per mouse, and $n = 3$ vehicle treated and 4 SRI treated. ***$P < 0.0001$, simple linear regression of slopes. Vehicle versus SRI and versus *Fire*$^{+/+}$, ***$P < 0.0001$; SRI versus *Fire*$^{+/+}$, $P = 0.0519$, Kruskal–Wallis with Dunn's multiple comparisons test. **n**, Myelin thickness versus axon diameter. $n = 100$ axons per mouse, and $n = 3$ vehicle treated and 4 SRI treated. ***$P < 0.0001$, simple linear regression of slopes. Vehicle versus SRI and versus *Fire*$^{+/+}$, ***$P < 0.0001$; SRI versus *Fire*$^{+/+}$, $P > 0.9999$, Kruskal–Wallis with Dunn's multiple comparisons test. Scale bars, 1 µm (**f**,**g**,**k**,**l**) or 25 µm (**b**).

Source data

We next asked whether stimulating TGFβR1 signalling in *Fire*$^{\Delta/\Delta}$ mice could rescue myelin health. In the absence of sufficient TGFβR1 expression by oligodendrocytes in *Fire*$^{\Delta/\Delta}$ mice, we used a small-molecule activator of downstream TGFβ signalling, SRI-011381 hydrochloride. This approach bypasses the need to stimulate the receptor by activating the SMAD2–SMAD3 pathway[25]. We administered SRI-011381 hydrochloride to *Fire*$^{\Delta/\Delta}$ mice from 2 months of age to observe the potential impact on the significant myelin pathology observed by 3 months of age (Fig. 6j). SRI-011381 had no influence on myelinated axon number (Extended Data Fig. 10f) but significantly reduced inner tongue thickness (Fig. 6k–m and Extended Data Fig. 10g) and myelin thickness (Fig. 6k,l,n and Extended Data Fig. 10h) compared with vehicle-treated *Fire*$^{\Delta/\Delta}$ mice. Following SRI-011381

hydrochloride treatment, these parameters overlapped with those in age-matched *Fire*$^{+/+}$ mice (Fig. 6m,n). This result suggested that disrupted TGFβR1 signalling is the primary mechanism by which myelin is dysregulated in *Fire*$^{\Delta/\Delta}$ mice. Altogether, these findings reveal the importance of the TGFβ1–TGFβR1 axis in microglia–oligodendrocyte communication for the regulation of myelin health.

Discussion

Here we identified the requirement for microglia in the maintenance of healthy CNS myelin. Our use of a new transgenic model in which microglia are lacking while other CNS macrophages are present revealed that microglia are not required for developmental oligodendrocyte maturation or myelin ensheathment. Rather, microglia preserve the structural integrity of myelin. We demonstrated the role of microglia in limiting hypermyelination and preventing demyelination of existing myelin sheaths in adulthood. Our results complement recent work implicating microglia in the regulation of myelin sheath number during myelin formation in embryonic development[26] and the inhibition of myelination of regenerated axons after optic nerve injury[27]. This raises the question of whether other macrophage populations, such as perivascular and/or peripheral macrophages, influence developmental myelination or whether other glial cell types, such as astrocytes, have a compensatory role in the absence of microglia. Our work indicates that a threshold number of microglia is needed to maintain myelin health, as even just a 50–60% decrease in white matter microglia in mice or humans is associated with loss of myelin integrity. Altogether, these findings suggest prudence in the current trials of CSF1R inhibitors to deplete microglia in cancer or neurological conditions, and warrant the monitoring of potential off-target effects on myelin health.

We associated structural changes in myelin in the absence of microglia with poor cognitive flexibility, along with impaired de novo myelination that normally underpins long-term memory consolidation[17,28]. This builds on previous work revealing that microglia dysregulation (in response to chemotherapy) is sufficient to disrupt myelin structure and cognitive function[9,13], as we have uncovered the requirement for healthy microglia in

the prevention of these pathologies. Given the close relationship between myelin structure and neuronal activity[29], our findings also raise the possibility of a role for microglia in influencing adaptive myelination to reinforce cognitive circuits. However, understanding the impact of the absence of microglia on neuronal activity and synaptic health requires further investigation. In addition, our work has important implications for understanding cellular networks that contribute to cognitive decline with ageing, in which there is prominent hypermyelination[3,12], demyelination and impaired new myelination[5,28], alongside microglia dysfunction[21,30,31]. This may also be relevant to dementia-associated neurodegenerative disease, given that in a mouse model of Alzheimer’s disease, there are opposite changes in gene modules primarily associated with microglia (and astrocytes) compared with those related to oligodendrocytes and myelination[32]. That study also identified that the oligodendrocyte-associated module is initially upregulated followed by a downregulation in microenvironments with the highest β-amyloid accumulation. Whether this represents initial hypermyelination followed by demyelination remains to be determined.

The worsening myelin pathology we observed with age in response to microglia depletion points to an increasing dependence on healthy microglia for myelin integrity with ageing. Microglia dysfunction may therefore initiate myelin damage with ageing and in neurodegenerative disorders. Previous work has implicated the peripheral immune system or primary oligodendrocyte dysregulation in inducing demyelination. Here we propose that the contribution of microglia now also needs to be considered. Our data suggest that hypermyelination may precede demyelination, which raises the question of whether this sequence of events underpins myelin damage in ageing and neurodegenerative disease. Notably, we identified parallels in dysregulated cellular profiles and molecular mechanisms between microglia-deficient mice and other neurological injury models. These results implicate microglia as crucial regulators of myelin pathology in these contexts. For instance, our data show that microglia normally suppress the appearance of a dysregulated oligodendrocyte state (expressing *Serpina3n* and *C4b*), similar to that recently documented in mouse models of demyelination, ageing and Alzheimer’s disease[33,34,35,36,37]. This implies that in pathological contexts, microglia dysregulation may permit these oligodendrocytes to appear. We have made data-mining of this (and other) oligodendrocyte populations in the

Fire$^{\Delta/\Delta}$ mouse model publicly available on the following website: https://annawilliams.shinyapps.io/shinyApp_oligos_VM/. The functional consequences of the appearance of these oligodendrocytes, and the molecular mechanisms involved in their function, have hitherto been unclear. Our association of these oligodendrocytes with an altered lipid profile (for example, increased cholesterol esters) is consistent with the hypermyelination in *Fire*$^{\Delta/\Delta}$ mice. This may shed light on potential pathological mechanisms in human neurological conditions in which cholesterol esters are increased, such as Huntington's disease and multiple sclerosis[38,39]. The finding that microglia regulate the lipid profile of mature myelinating oligodendrocytes complements their recently discovered role in promoting oligodendrocyte progenitor maturation during myelin regeneration through the supply of a cholesterol pathway intermediate[40]. Our pathway analysis of these oligodendrocytes identified a dysregulated TGFβ–TGFβR1 axis, which we showed is a mediator of myelin pathology. Transcriptomic analyses of ALSP brain samples has indicated dysregulation of the TGFβ pathway specifically in the white matter[20], and an upregulation of *TGFB1* in the least affected white matter region[41]. Signalling downstream of TGFβ receptors is also reduced in ageing and neurodegenerative disease[42,43].

We previously discovered that a subset of microglia expressing a TGFβ superfamily member, activin-A, regulates remyelination efficiency[44,45]. Taken together with the transcriptomic heterogeneity of microglia during development, homeostasis, demyelination, remyelination and ageing[30,44,46,47,48,49,50], we are now poised to ask whether specific microglia states are required to regulate myelin growth and integrity. Recently, studies have identified microglia states associated with white matter, with roles in phagocytosis of dying cells in development[50] or myelin debris in ageing[30], and we found that a shift in functional microglial states underpins their capacity to support remyelination[44,46]. Whether altered heterogeneity with ageing and disease is associated with a loss in supportive microglia states that then contributes to progressive myelin pathology needs to be investigated. Altogether, our study uncovered the role of microglia in preserving myelin health and integrity in adulthood, and highlights microglia as key therapeutic targets in the context of disrupted myelin in ageing and disease.

Methods

Animals

All experiments were performed under project licences approved by the UK Home Office and issued under the Animals (Scientific Procedures) Act. This study used *Csf1r*$^{Fire\Delta/\Delta}$mice, wild-type controls from *Csf1r*$^{Fire\Delta/+}$ crossings, *Plp*^{creERT} mice (Jackson Laboratories) and *Tgfbr1*$^{fl/fl}$ mice (provided by S. Karlsson, Lund University). Recombination was induced in *Plp*^{creERT};*Tgfbr1*$^{fl/fl}$ mice through the administration of 4-hydroxytamoxifen (100 mg kg^{-1}, intraperitoneally; Sigma-Aldrich) dissolved in ethanol and corn oil (1:9) mixture for 5 consecutive days from P14 to P18, then killed at P28. All animals were housed at a maximum number of 6 animals per cage in a 12 h light–dark cycle with unrestricted access to food and water. For animal experiments, the sample size was determined by power analysis calculated by two-sided 95% confidence interval through the normal approximation method using OpenEpi software (Openepi.com), and reached >80% power for all experiments. Both males and females were used throughout the study, except for open-field experiments, for which only male mice were used. Animals were randomized to time points analysed. ARRIVE2 guidelines were followed in providing details of experiments, quantifications and reporting.

Genotyping

Genomic DNA was extracted from ear biopsy tissue using a Wizard SV genomic purification system (Promega) according to the manufacturer's instructions. *Csf1r*$^{Fire\Delta/\Delta}$ mice were genotyped using PCR strategies as previously described[10]. Genotyping of *Plp*^{creERT};*Tgfbr1*$^{fl/fl}$ mice was performed with genomic DNA extracted from the tail, as previously described for *Plp*^{creERT} mice[51] and *Tgfbr1*$^{fl/fl}$mice[52].

Immunofluorescence staining of mouse tissue

Mice were intracardially perfused with 4% paraformaldehyde (PFA; w/v; Sigma), and brains were post-fixed overnight and cryoprotected in sucrose before embedding in OCT (Tissue-Tech) and storage at −80 °C. Cryosections (10 μm) were air-dried, permeabilized and blocked for 1 h with 5% normal horse serum (Gibco) and 0.3% Triton-X-100 (Fisher Scientific) in PBS. For myelin protein staining, sections were permeabilized in methanol at −20 °C for 10 min. For EdU visualization, an AlexaFluor-555 Click-iT EdU Cell Proliferation Assay kit (Invitrogen) was applied before immunostaining. Sections were permeabilized with 0.5% Triton X-100 in PBS for 20 min at room temperature (20–25 °C) then incubated in Click-iT reaction cocktail in the dark at room temperature for 30 min and then washed in PBS. Heat-induced antigen retrieval was performed before primary antibody application. Primary antibodies were applied overnight at 4 °C in a humid chamber and included the following: MBP (AbD Serotec, 1:250; MCA409S, clone 12); MAG (Millipore, 1:100; MAB1567, clone 513); MOG (Millipore, 1:100; MAB5680, clone 8-18C5); CNPase (Sigma-Aldrich, 1:100; AMAB91072, clone CL2887); TMEM119 (Abcam, 1:100; ab209064, clone 28-3); IBA1 (Abcam, 1:500; ab5076); CD206 (Abcam, 1:100; ab64693); CD31 (R&D Systems, 1:100; AF3628); LYVE1 (Abcam, 1:100; ab14917); OLIG2 (Millipore, 1:100; AB9610, clone 211F1.1); APC/CC1 (Abcam, 1:100; ab16794, clone CC1); SOX9 (Millipore, 1:500; AB5535); GFAP (Cambridge Bioscience, 1:500; 829401); neurofilament-H (BioLegend, 1:100,000; Covance, PCK-592P); and PLP (Abcam, 1:100; ab28486). For SERPINA3N immunostaining, 6 μm-thick formalin-fixed paraffin-embedded sections were deparaffinized, rehydrated and then placed in a water bath at 85 °C for 30 min in a citrate-based antigen unmasking solution (H-3300-250, Vector Laboratories). Sections were subsequently rinsed in PBS and blocked for 1 h using PBS with 10% donkey serum (D9663, Sigma Aldrich) and 0.2% Triton X-100 (T8787, Sigma Aldrich). Sections were then incubated overnight at 4 °C in anti-SERPINA3N (R&D Systems, 1:100; AF4709) and anti-OLIG2 (Millipore, 1:500; AB9610) in PBS with 5% donkey serum and 0.1% Triton X-100, then washed three times in PBS with 0.05% Tween-20 (P1379, Sigma Aldrich). Fluorescently conjugated secondary antibodies were applied for 1–2 h at room temperature in a humid chamber (1:500, Life Technologies–Molecular Probes). Following counterstaining with Hoechst or DAPI, slides were coverslipped with Fluoromount-G (Southern Biotech). For TGFβR1 immunostaining of *Fire*$^{\Delta/\Delta}$ mice, following a wash in TBS with 0.001%

Triton X-100 (Sigma), sections were microwaved in antigen unmasking solution (pH 6 citrate buffer, Vector Laboratories) for 10 min, then heated at 60 °C for 30 min. After cooling, sections were washed once with TBS and 0.001% Triton X-100, and endogenous phosphatase and peroxidase activity was blocked with Bloxall (Vector) for 10 min. Blocking was performed for 1 h with 10% heat-inactivated horse serum (Gibco) and 0.5% Triton X-100 in TBS. Primary antibody diluted in blocking solution was applied overnight in a humid chamber at 4 °C. Antibodies used included TGFβR1 (Abcam, 1:100; ab31013) and OLIG2 (Millipore, 1:100; MABN50). Following three washes in TBS with 0.001% Triton X-100, peroxidase-conjugated secondary antibody (Vector) was applied for 1 h at room temperature in a humid chamber. Following further washes, sections were developed using Opal 520 (Akoya) at 1:100 in Plus Amplification Diluent (Akoya) for 10 min in a humid chamber. Slides were washed, and residual peroxidase activity was quenched by applying Bloxall (Vector) for 10 min. For co-staining, another primary antibody was then applied and developed using peroxidase-conjugated secondary antibody and Opal 650 (Akoya) at 1:100 in Plus Amplification Diluent as described above. Following washes in TBS, the sections were counterstained with Hoechst (1:10,000) and mounted with Fluoromount-G (Invitrogen).

Sections were imaged on a Leica SPE or Zeiss LSM 510 confocal microscope. Cell counts were calculated from a measured area based on assumption of circularity using Fiji/ImageJ (Fiji.sc), with three regions of interest quantified per section. Colocalization analysis of SOX9 and GFAP was performed using Imaris software v.9.7.

Flow cytometry

An enzyme-free brain dissociation protocol was used to gather a myeloid cell-enriched cell suspension to explore CSF1R (CD115) surface protein levels. After transcardially perfusing 10–11-week-old female mice with ice-cold PBS, brains were dissected and minced with a 22A scalpel in HBSS (without Ca^{2+} and Mg^{2+}; 14175-053, Gibco) with 25 mM HEPES (10041703, Fisher Scientific). Brains were then homogenized using a Dounce homogenizer (D9938, Kimble) in HBSS (without Ca^{2+} and Mg^{2+}) with 25

mM HEPES. Brain homogenates were separated using a 35% Percoll gradient, with centrifugation at 800*g* for 20 min at 4 °C (with no brake). Cell pellets were collected and washed in PBS (without Ca^{2+} and Mg^{2+}; 14190-094, Gibco) with 0.1% low endotoxin BSA (A8806, Sigma Aldrich). Fc receptors were blocked (1:100; 101302, BioLegend) for 15 min at 4 °C on a shaker. Cells were then stained with primary antibodies directed against CD11b (PE; 1:200; 101207; BioLegend, clone M1/70), CD45 (PECy7; 1:200; 103114; BioLegend, clone 30-F11) and CD115 (APC; 1:200; 135510; BioLegend, clone AFS98) for 30 min at 4 °C on a shaker. Samples were then washed and resuspended in PBS (without Ca^{2+} and Mg^{2+}) with 0.1% low endotoxin BSA. Single-fluorochrome stained beads, unstained samples and fluorescence minus one samples were used as controls. DAPI was used for cell viability gating. Data were acquired using a BD LSRFortessa flow cytometer. FCS express 7 was used for post-acquisition data analysis.

Electron microscopy of mouse tissue

Mice were intracardially perfused with 4% PFA (w/v) and 2% glutaraldehyde (v/v; TAAB Laboratories) in 0.1 M phosphate buffer. Tissue was post-fixed overnight at 4 °C and transferred to 1% glutaraldehyde (v/v) until embedding. Tissue sections (1 mm) were post-fixed in 1% osmium tetroxide and dehydrated before processing into araldite resin blocks. Next, 1 µm microtome-cut sections were stained with a 1% toluidine blue/2% sodium borate solution before bright-field imaging using a Zeiss Axio microscope. Ultrathin sections (60 nm) were cut from corpus callosum samples, stained with uranyl acetate and lead citrate, and grids imaged on a JEOL transmission electron microscope. Axon diameter, myelin and inner tongue thickness were calculated from a measured area based on assumption of circularity using Fiji/ImageJ (Fiji.sc) (diameter = $2 \times \sqrt{[area/\pi]}$), with 100–200 axons per animal analysed.

Behavioural testing

Experimenters were blinded to genotype during behavioural testing and data analyses. All experiments were performed in a behaviour testing room maintained at a constant temperature of 20 °C. The open-field test was

performed on male mice at 4–8 weeks and 11–13 months of age to assess locomotor activity and anxiety-associated behaviours. Handling was carried out 3–4 days before testing. Mice were placed in the open field (47 × 47 cm) to freely explore the arena for 10 min. Equipment was cleaned with 70% ethanol between each test to remove odours. The total ambulatory distance travelled (in metres) and the time spent in the edges (9 cm from the wall) and centre (29 × 29 cm) were automatically quantified using the video tracking software Any-Maze (Stoelting Europe, v.6.3). The Barnes maze test was performed in adult mice 2–4 months of age to assess spatial learning, memory and cognitive flexibility. The maze consisted of one white circular platform with 20 circular holes around the outside edge, 91.5 cm in diameter and 115 cm in height (San Diego Instruments). A dark escape chamber was attached to one of the holes, and the location of the escape chamber remained constant for each mouse but was shifted 90° clockwise between consecutive mice to avoid carryover of olfactory cues. Lamps and overhead lights (450 lux) were used to light the maze. Once the trial started, an aversive white noise stimulus at 85 dB was played until the mouse entered the escape chamber. Visual cues were present on the curtains and walls around the maze. Animals were retained within a white holding cylinder (diameter of 10.5 cm) at the beginning of each trial. The maze and escape chamber were cleaned with ethanol between each trial to avoid carryover of olfactory cues between animals. All trials were recorded using the video-based automated tracking software Any-Maze (Stoelting Europe, v.4.99). Before testing, mice were handled for 3–5 min per day for 6–7 days by the experimenter. Animals were brought into the testing room and placed in the holding cylinder to acclimate to the testing environment for 10 s for 2 days before habituation. Mice were habituated to the maze and escape chamber 1 day before the start of the learning phase, whereby each mouse was placed in the holding cylinder for 10 s then allowed to freely explore the maze with no aversive stimuli for 3 min. Mice were then guided to the escape chamber and retained inside for 1 min. During the learning phase (T1–T6), mice were trained to locate the escape chamber over 6 consecutive days with 2 trials per day (1 h inter-trial interval); data per mouse were averaged per day. If the mouse failed to enter the escape chamber during the 3-min trial period, the experimenter guided it to the chamber. Spatial learning was assessed by the total time taken to locate the escape chamber in each trial (primary latency; defined by the head entering the chamber), and spatial working memory was assessed by the

number of errors made before locating the escape chamber (primary errors; defined by the nose deliberately entering a hole with some extension of the head and neck or hindpaws). The total distance travelled and speed during the trials were additionally measured. Exclusion criteria were defined before data analysis as follows: mice must enter a minimum of three quadrants of the maze within two of the first five trials and must enter the escape chamber during the first three trials. One wild-type and one knockout mouse were excluded from analysis owing to refusal to enter the chamber. At 1 h and 3 days following the last trial, probe tests were performed whereby the mouse was allowed to explore the maze for 1 min with the escape chamber removed. The time spent in the target quadrant of the maze and the number of nose pokes in each hole were recorded to assess memory of the escape chamber location. To assess cognitive flexibility, mice underwent the reversal learning phase (R1–R3), whereby the escape box was moved to 180° from the original location, and measurements were taken as described above. A probe test was also performed 3 days after the final reversal trial. The median age of mice assessed that were trained and untrained were comparable between genotypes at the time of euthanasia for immunofluorescence or electron microscopy analysis: untrained $\mathit{Fire}^{+/+}$ and $\mathit{Fire}^{\Delta/\Delta}$ mice were 118 days old, trained $\mathit{Fire}^{+/+}$ mice were 119 days old and trained $\mathit{Fire}^{\Delta/\Delta}$ mice were 120 days old.

EdU incorporation

EdU was dissolved in the drinking water at 0.2 mg ml^{-1} for a period of 14 days from the end of trial day 1 until the end of the experiment. The water was exchanged every other day, and intake was monitored to assess whether consistent volumes were consumed.

Microglia depletion in adulthood

The CSF1R inhibitor PLX5622 (Chemgood, C-1521) was formulated into chow at a concentration of 1,200 ppm (Research Diets, D11100404i) and fed to 2-month-old and 5-month-old wild-type ($\mathit{Fire}^{+/+}$) mice for 1 month and euthanized as described above.

SRI-011381 hydrochloride administration

The TGFβ signalling agonist SRI-011381 hydrochloride (HY-100347A, Cambridge Bioscience/MedChem Express), dissolved in PBS containing DMSO (10%) and PEG300 (40%), was injected intraperitoneally into *Fire*$^{\Delta/\Delta}$ mice at a dose of 30 mg kg^{-1} 3 times per week for 3 weeks (9 injections in total) and killed as described above.

Measuring TGFβ1 levels by ELISA

Following perfusion with PBS, the corpus callosum was dissected from 2 mm coronal sections of *Fire*$^{\Delta/\Delta}$ and wild-type brains and snap-frozen in liquid nitrogen. Samples were homogenized in RIPA buffer (Millipore, 20–188) containing phosphatase and protease inhibitors (Sigma-Aldrich, 4906845001 and 11836170001). Corpus callosum lysate samples were activated and TGFβ1 protein levels were measured by ELISA (BioLegend, 436707) according to the manufacturer's instructions for serum and plasma samples at a final dilution of 1:10. BCA assays were performed to measure total protein according to the manufacturer's instructions (Thermo Fisher Scientific, 23225), and values were normalized to this for each sample.

Brain dissociation and cell sorting for single-cell RNA sequencing

Brains were collected from 6–7-week-old female mice at the same time of day for each animal. Mice were culled by cervical dislocation and brains were dissected, with the olfactory bulbs and cerebellum removed. Hippocampi from both hemispheres and the remainder of the left hemisphere (without the hippocampus) were collected in ice-cold HBSS (without Ca^{2+} and Mg^{2+}; 14175-053, Gibco) with 5% trehalose (T0167, Sigma Aldrich) and 30 µM actinomycin D (A1410, Sigma Aldrich) and were finely minced using a 22A scalpel. Brains were digested using the an Adult Brain Dissociation kit (130-107-677, Miltenyi Biotec) with the following modifications: (1) tissues were dissociated as described in the "manual dissociation" section of the Neural Tissue Dissociation kit protocol (130-092-628, Miltenyi Biotec); (2)

enzymatic digestions were performed at 35 °C; (3) half the concentration of enzyme P was used; (4) actinomycin D was used to limit dissociation-induced transcriptional changes; (5) 5% trehalose was added in all buffers to increase cellular viability; (6) cell clusters were removed by filtration through pre-moistened 70 µm (352350, Falcon) and 40 µm (352340; Falcon) cell strainers; (7) erythrocyte and myelin debris removal steps were omitted during dissociation steps; and (8) all centrifugations were performed at 200*g* at 4 °C. After dissociation, cells were collected in PBS with 0.2% BSA before being sorted on a Sony SH800 cell sorter. Gates were chosen based on forward and side scatter to exclude myelin debris, erythrocytes and doublets. Non-viable cells were excluded based on DRAQ7 and/or DAPI staining (DRAQ7^{high} and/or DAPI^{high} cells were classified as non-viable). After confirming the viability of cells after sorting, using trypan blue and a haemocytometer, single cells were processed through the Chromium Single Cell Platform using a Chromium Next GEM Single Cell 3′ GEM Library and Gel Bead kit (v.3.1 chemistry, PN-1000121, 10x genomics) and a Chromium Next GEM Chip G kit (PN-1000120) and processed following the manufacturer’s instructions. Libraries were sequenced using a NovaSeq 6000 sequencing system (PE150 (HiSeq), Illumina).

Pre-processing of sequencing data and single-cell RNA sequencing analysis

Alignment to the reference genome, feature counting and cell calling were performed following the 10x Genomics CellRanger (v.5.0.0) pipeline, using the default mm10 genome supplied by 10x Genomics (https://cf.10xgenomics.com/supp/cell-exp/refdata-gex-mm10-2020-A.tar.gz). From the output, the filtered matrices were used for downstream analyses. Pre-processing was performed on the University of Edinburgh’s compute cluster Eddie. The analysis was performed with R v.4.1.1. Full details to replicate the analysis pipelines described below can be found in code scripts available on GitHub (https://github.com/Anna-Williams/Veronique-Firemice). SingleCellEXperiment v.1.14.1 was used to handle the single-cell experiment objects in R. Cells were filtered using dataset-specific parameters on the basis of genes and unique molecular identifiers (UMIs) per cell, the ratio between these two parameters and the

percentage of mitochondrial gene reads per cell. Thresholds were computed with the isOutlier function from scater (v.1.20.1), as batch 6 was of poorer quality than the other batches, with outlier values, the subset argument was used. Only genes that were detected in at least two cells were kept. Using scran (v.1.20.1), the data were normalized by deconvolution, and the top 15% highly variable genes were selected. Following principal component analysis (PCA), 25 principal components (PCs) were kept for downstream analysis (cut-off selected by examination of an Elbowplot). Nonlinear dimensional dimension representation and t-distributed stochastic neighbour embedding (t-SNE) and gene expression variance explained by batch (computed with scater) revealed the need for batch correction. Batch correction was performed using mutual nearest neighbours with fastMNN, batchelor (v.1.8.0). Finally, a graph-based clustering approach was used to cluster the cells using the clusterCells function from scran, with $k = 60$. Clusters with the highest expression of oligodendrocyte markers (*Plp1*, *Mog*, *Mag* and *Mbp*) and that did not express other cell type markers (for example, astrocyte, oligodendrocyte progenitor cell or microglia markers) were subsetted to be analysed separately. Cell and gene quality control were further adjusted, setting a stricter minimum UMI count threshold (5,000 UMIs) and maximum percentage of mitochondrial gene reads per cell (10%). A small cluster of cells of lower quality based on the new thresholds was also excluded from the analysis (Extended Data Fig. 10). Ultimately, we included a total of 19,506 genes and 13,583 cells. The normalization, feature selection, dimensional reduction and batch correction were repeated with the subset dataset as described above. Clustering was performed at different resolutions, and after examination with clustree (v.0.4.4), $k = 100$ was selected and then merged into four clusters (Fig. 6a and Extended Data Fig. 10). Differential gene expression between cluster 1 (specific to the *Fire*$^{\Delta/\Delta}$ mice) and the mean expression of all other cells was performed using FindMarkers from Seurat (v.4.1.0) (Supplementary Table 1). ShinyCell v.2.1.0 was used to produce an interactive application, and org.Mm.eg.db v.3.13.0 was used to annotate genes. ggplot2 v.3.3.5 was used to perform custom plots, here v.1.0.1 was used to ensure reproducible paths, and Matrix v.1.3.4 was used for handling sparse matrices.

Lipidomics

Corpus callosum samples from $Fire^{+/+}$ and $Fire^{\Delta/\Delta}$ mice were diluted in 700 μl of PBS with 800 μl 1 N HCl:CH_3OH 1:8 (v/v), 900 μl $ChCl_3$ and 200 μg ml^{-1} of the antioxidant 2,6-*di*-tert-butyl-4-methylphenol (BHT; Sigma-Aldrich). Splash Lipidomix Mass Spec standard (3 μl; Avanti Polar Lipids) was added into the extract mix. The organic fraction was evaporated at room temperature using a Savant Speedvac spd111v (Thermo Fisher), and the remaining lipid pellet was reconstituted in 100% ethanol. Lipid species were analysed by liquid chromatography electrospray ionization tandem mass spectrometry (LC–ESI–MS/MS) on a Nexera X2 UHPLC system (Shimadzu) coupled with hybrid triple quadrupole/linear ion trap mass spectrometer (6500+ QTRAP system; AB SCIEX). Chromatographic separation was performed on a XBridge amide column (150 × 4.6 mm, 3 × 5 μm; Waters) maintained at 35 °C using mobile phase A (1 mM ammonium acetate in water and acetonitrile 5:95 (v/v)) and mobile phase B (1 mM ammonium acetate in water and acetonitrile 50:50 (v/v)) in the following gradient: 0–6 min: 0% B; 6% B; 6–10 min: 6% B; 25% B; 10–11 min: 25% B; 98% B; 11–13 min: 98% B; 100% B; 13–19 min: 100% B; 19–24 min: 0% B. The flow rate was set at 0.7 ml min^{-1}, which was increased to 1.5 ml min^{-1} from 13 min onwards. Sphingomyelin and cholesteryl esters were measured in positive ion mode with a precursor scan of 184.1, 369.4. Triglycerides, diglycerides and monoglycerides were measured in positive ion mode with a neutral loss scan for one of the fatty acyl moieties. Phosphatidylcholine, phosphatidylethanolamine, phosphatidylglycerol, phosphatidylinositides and phosphatidylserines were measured in negative ion mode by fatty acyl fragment ions. Lipid quantification was performed by scheduled multiple reactions monitoring, with the transitions based on the neutral losses or the typical product ions as described above. The instrument parameters were set as follows: curtain gas = 35 psi; collision gas = 8 a.u. (medium); IonSpray 16 voltage = 5,500 V and −4,500 V; temperature = 550 °C; ion source gas 1 = 50 psi; ion source gas 2 = 60 psi; declustering potential = 60 V and −80 V; entrance potential = 10 V and −10 V; and collision cell exit potential = 15 V and −15 V. Peak integration was performed using the Multiquant TM software v.3.0.3. Lipid species signals were corrected for isotopic contributions (calculated using Python Molmass 2019.1.1) and were quantified on the basis of internal standard signals and adhere to the guidelines of the Lipidomics Standards Initiative.

Immunofluorescence staining of human tissue

Post-mortem tissue from individuals with ALSP and from unaffected individuals who had died of non-neurological causes (Extended Data Table 1) were obtained with full ethical approval from the Queen Square Brain Bank for Neurological Disorders, UCL Queen Square Institute of Neurology, and their use was in accord with the terms of the informed consents for use of donor tissue and information. Ethical approval was granted to the Queen Square Brain Bank for the use of tissue by the National Health Service Health Research Authority through the London Central Research Ethics Committee. Diagnosis of ALSP was confirmed on the basis of mutations in *CSF1R* and by neuropathological means, and clinical history was provided by Z. Jaunmuktane (University College London). Formalin-fixed paraffin embedded tissue blocks were cut to 10 μm thickness. Sections were placed in the oven at 60 °C for 10 min and deparaffinized by a series of washes in HistoClear (2× 10 min) and ethanol (100% twice, 95%, 70%, 50%; 5 min each). Following washes in PBS, slides were placed in a pressure cooker in Vector unmasking solution for 20 min, cooled, washed once and blocked for 1 h with 5% normal horse serum (Gibco) and 0.3% Triton X-100 (Fisher Scientific) in PBS. Tissue was incubated with primary antibodies in a humid chamber overnight. Sections were then washed in PBS, and fluorescently conjugated secondary antibodies were applied for 2 h at room temperature in a humid chamber (1:500, Life Technologies–Molecular Probes). Following washes in PBS and then water, the sections were counterstained with Hoechst and washed with distilled water for 20 min. Slides were then coverslipped with Fluoromount-G (Southern Biotech). Primary antibodies included IBA1 (Abcam, 1:500; ab5076) and LYVE1 (Abcam, 1:100; ab14917). Entire tissue sections were imaged using a Zeiss AxioScan Z.1 SlideScanner and Zeiss Zen2 software (blue edition). Three fields of 150 × 150 μm were counted per region of interest in each case and counts were multiplied to determine the density of immunopositive cells per mm^2.

Electron microscopy of human tissue

ALSP tissue was acquired from the Department of Neuropathology at Charité-Universitätsmedizin Berlin, and tissue from unaffected individuals

were obtained from the Medical Research Council Edinburgh Brain and Tissue Bank, approved by the respective ethical review boards. ALSP tissue samples of the white matter were fixed in 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer for 48 h at 4 °C. Samples were post-fixed in 1% osmium tetroxide in 0.05 M sodium cacodylate buffer for 3 h, dehydrated in graded acetone series including en bloc staining with 1% uranyl acetate and 0.1% phosphotungstic acid in the 70% acetone step for 60 min and embedded in araldite resin. For tissue from unaffected individuals, ultrathin sections were stained with uranyl acetate and lead citrate and imaged with a Zeiss P902 electron microscope. For ALSP tissue, ultrathin sections with virtual absence of limiting artefacts were prepared and entirely digitized using a Zeiss Gemini 300 scanning electron microscope with a scanning transmission electron microscopy. In brief, we used 29 kV acceleration voltage, 5 nm pixel size and 1.5 µs beam dwell time for digitization and Fiji/TrakEM2 for stitching to allow for in-depth analysis with QuPath 0.3.0.

Statistics and reproducibility

All manual cell counts were performed in a blinded manner. Data are presented as the mean ± s.e.m. All micrographs are representative images of the result, which were independently repeated a minimum of three times; exact *n* values are presented in the corresponding graphs of quantification. Before statistical testing, data were assessed for normality of distribution using Shapiro–Wilk test. Statistical tests included two-tailed Student's *t*-test for normally distributed data or Mann–Whitney test for nonparametric data, two-way analysis of variance (ANOVA) with Sidak's multiple comparisons test for comparing more than two groups, and a one-sample *t*-test for comparison of $\log_2$(fold change) to a value of 0. For proportion graphs, one-way ANOVA with Tukey's multiple comparisons test was used. For Barnes maze analysis of primary latency and errors, repeated measures two-way ANOVA with Sidak's multiple comparisons test was used. Before testing for inter-group differences in the probe test data, each genotype was tested against chance (25%) using a one sample *t*-test. Slopes of myelin and inner tongue thickness versus axon diameter were compared using simple linear regression analysis, and for comparison of more than two groups, Kruskal–Wallis with Dunn's multiple comparisons test. A P value of <0.05 was

considered significant. Data handling and statistical processing were performed using Microsoft Excel and GraphPad Prism Software.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw single-cell RNA sequence datasets have been deposited in the Gene Expression Omnibus with accession code GSE215440. Analysed oligodendrocyte sequencing data are perusable on the following shiny application: https://annawilliams.shinyapps.io/shinyApp_oligos_VM/. Full details to replicate the analysis pipelines can be found in code scripts available on GitHub (https://github.com/Anna-Williams/Veronique-Firemice). Alignment to the reference genome, feature counting and cell calling were performed following the 10x Genomics CellRanger (v.5.0.0) pipeline, using the default mm10 genome supplied by 10x Genomics (https://cf.10xgenomics.com/supp/cell-exp/refdata-gex-mm10-2020-A.tar.gz). Source data are provided with this paper.

Code availability

Full details to replicate the analysis pipelines described below can be found in code scripts available on GitHub (https://github.com/Anna-Williams/Veronique-Firemice).

References

1. McNamara, N. B. & Miron, V. E. Microglia in developing white matter and perinatal brain injury. *Neurosci. Lett.* **714**, 134539 (2020).

2. Romanelli, E. et al. Myelinosome formation represents an early stage of oligodendrocyte damage in multiple sclerosis and its animal model. *Nat. Commun.* **7**, 13275 (2016).

3. Peters, A. The effects of normal aging on myelin and nerve fibers: a review. *J Neurocytol.* **31**, 581–593 (2002).

4. Williamson, J. M. & Lyons, D. A. Myelin dynamics throughout life: an ever-changing landscape? *Front. Cell Neurosci.* **12**, 424 (2018).

5. Safaiyan, S. et al. Age-related myelin degradation burdens the clearance function of microglia during aging. *Nat. Neurosci.* **19**, 995–998 (2016).

6. Inagawa, K., Watanabe, S., Tsukada, Y. & Mikoshiba, K. The role of myelination in learning performance observed in two strains of myelin-deficient mutant mice (shiverer and mld). *Behav. Neural. Biol* . **50**, 184–192 (1988).

7. Peters, A., Moss, M. B. & Sethares, C. Effects of aging on myelinated nerve fibers in monkey primary visual cortex. *J. Comp. Neurol.* **419**, 364–376 (2000).

8. Ritchie, S. J. et al. Coupled changes in brain white matter microstructure and fluid intelligence in later life. *J. Neurosci.* **35**, 8672–8682 (2015).

9. Gibson, E. M. et al. Methotrexate chemotherapy induces persistent tri-glial dysregulation that underlies chemotherapy-related cognitive impairment. *Cell* **176**, 43–55.e13 (2019).

10. Rojo, R. et al. Deletion of a *Csf1r* enhancer selectively impacts CSF1R expression and development of tissue macrophage populations. *Nat. Commun.* **10**, 3215 (2019).

11. Munro, D. A. D. et al. CNS macrophages differentially rely on an intronic *Csf1r* enhancer for their development. *Development* **147**, dev194449 (2020).

12. Phillips, K. A. et al. Age-related changes in myelin of axons of the corpus callosum and cognitive decline in common marmosets. *Am. J. Primatol.* **81**, e22949 (2019).

13. Geraghty, A. C. et al. Loss of adaptive myelination contributes to

methotrexate chemotherapy-related cognitive impairment. *Neuron* **103**, 250–265.e8 (2019).

14. Hiramoto, T. et al. *Tbx1*, a gene encoded in 22q11.2 copy number variant, is a link between alterations in fimbria myelination and cognitive speed in mice. *Mol. Psychiatry* **27**, 929–938 (2022).

15. Silva, A. I. et al. *Cyfip1* haploinsufficient rats show white matter changes, myelin thinning, abnormal oligodendrocytes and behavioural inflexibility. *Nat. Commun.* **10**, 3455 (2019).

16. McKenzie, I. A. et al. Motor skill learning requires active central myelination. *Science* **346**, 318–322 (2014).

17. Steadman, P. E. et al. Disruption of oligodendrogenesis impairs memory consolidation in adult mice. *Neuron* **105**, 150–164.e6 (2020).

18. Oosterhof, N. et al. Colony-stimulating factor 1 receptor (CSF1R) regulates microglia density and distribution, but not microglia differentiation in vivo. *Cell Rep.* **24**, 1203–1217.e6 (2018).

19. Saher, G. et al. High cholesterol level is essential for myelin membrane growth. *Nat. Neurosci.* **8**, 468–475 (2005).

20. Kempthorne, L. et al. Loss of homeostatic microglial phenotype in CSF1R-related leukoencephalopathy. *Acta Neuropathol. Commun.* **8**, 72 (2020).

21. Galatro, T. F. et al. Transcriptomic analysis of purified human cortical microglia reveals age-associated changes. *Nat. Neurosci.* **20**, 1162–1171 (2017).

22. Bottcher, C. et al. Human microglia regional heterogeneity and phenotypes determined by multiplexed single-cell mass cytometry. *Nat. Neurosci.* **22**, 78–90 (2019).

23. Michael, D. R., Salter, R. C. & Ramji, D. P. TGF-β inhibits the uptake of modified low density lipoprotein by human macrophages through a

Smad-dependent pathway: a dominant role for Smad-2. *Biochim. Biophys. Acta* **1822**, 1608–1616 (2012).

24. Butovsky, O. et al. Identification of a unique TGF-β-dependent molecular and functional signature in microglia. *Nat. Neurosci.* **17**, 131–143 (2014).

25. Wu, Q. et al. Astrocytic YAP protects the optic nerve and retina in an experimental autoimmune encephalomyelitis model through TGF-β signaling. *Theranostics* **11**, 8480–8499 (2021).

26. Hughes, A. N. & Appel, B. Microglia phagocytose myelin sheaths to modify developmental myelination. *Nat. Neurosci.* **23**, 1055–1066 (2020).

27. Wang, J. et al. Robust myelination of regenerated axons induced by combined manipulations of GPR17 and microglia. *Neuron* **108**, 876–886.e4 (2020).

28. Wang, F. et al. Myelin degeneration and diminished myelin renewal contribute to age-related deficits in memory. *Nat. Neurosci.* **23**, 481–486 (2020).

29. Gibson, E. M. et al. Neuronal activity promotes oligodendrogenesis and adaptive myelination in the mammalian brain. *Science* **344**, 1252304 (2014).

30. Safaiyan, S. et al. White matter aging drives microglial diversity. *Neuron* **109**, 1100–1117.e10 (2021).

31. Olah, M. et al. A transcriptomic atlas of aged human microglia. *Nat. Commun.* **9**, 539 (2018).

32. Chen, W. T. et al. Spatial transcriptomics and in situ sequencing to study Alzheimer's disease. *Cell* **182**, 976–991.e19 (2020).

33. Shen, K. et al. Multiple sclerosis risk gene *Mertk* is required for microglial activation and subsequent remyelination. *Cell Rep.* **34**,

108835 (2021).

34. Falcao, A. M. et al. Disease-specific oligodendrocyte lineage cells arise in multiple sclerosis. *Nat. Med.* **24**, 1837–1844 (2018).

35. Zhou, Y. et al. Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat. Med.* **26**, 131–142 (2020).

36. Dulken, B. W. et al. Single-cell analysis reveals T cell infiltration in old neurogenic niches. *Nature* **571**, 205–210 (2019).

37. Kenigsbuch, M. et al. A shared disease-associated oligodendrocyte signature among multiple CNS pathologies. *Nat. Neurosci.* **25**, 876–886 (2022).

38. Phillips, G. R. et al. Cholesteryl ester levels are elevated in the caudate and putamen of Huntington's disease patients. *Sci Rep.* **10**, 20314 (2020).

39. Pitkanen, A. S., Halonen, T. O., Kilpelainen, H. O. & Riekkinen, P. J. Cholesterol esterase activity in cerebrospinal fluid of multiple sclerosis patients. *J. Neurol. Sci.* **74**, 45–53 (1986).

40. Berghoff, S. A. et al. Microglia facilitate repair of demyelinated lesions via post-squalene sterol synthesis. *Nat. Neurosci.* **24**, 47–60 (2021).

41. Berdowski, W. M. et al. Dominant-acting CSF1R variants cause microglial depletion and altered astrocytic phenotype in zebrafish and adult-onset leukodystrophy. *Acta Neuropathol.* **144**, 211–239 (2022).

42. Colangelo, V. et al. Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: transcription and neurotrophic factor down-regulation and up-regulation of apoptotic and pro-inflammatory signaling. *J. Neurosci. Res.* **70**, 462–473 (2002).

43. Tichauer, J. E. et al. Age-dependent changes on TGFβ1 Smad3 pathway modify the pattern of microglial cell activation. *Brain Behav. Immun.*

37, 187–196 (2014).

44. Miron, V. E. et al. M2 microglia and macrophages drive oligodendrocyte differentiation during CNS remyelination. *Nat. Neurosci.* **16**, 1211–1218 (2013).

45. Dillenburg, A. et al. Activin receptors regulate the oligodendrocyte lineage in health and disease. *Acta Neuropathol.* **135**, 887–906 (2018).

46. Lloyd, A. F. et al. Central nervous system regeneration is driven by microglia necroptosis and repopulation. *Nat. Neurosci.* **22**, 1046–1052 (2019).

47. Hammond, T. R. et al. Single-cell RNA sequencing of microglia throughout the mouse lifespan and in the injured brain reveals complex cell-state changes. *Immunity* **50**, 253–271.e6 (2019).

48. Plemel, J. R. et al. Microglia response following acute demyelination is heterogeneous and limits infiltrating macrophage dispersion. *Sci. Adv.* **6**, eaay6324 (2020).

49. Masuda, T. et al. Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. *Nature* **566**, 388–392 (2019).

50. Li, Q. et al. Developmental heterogeneity of microglia and brain myeloid cells revealed by deep single-cell RNA sequencing. *Neuron* **101**, 207–223.e10 (2019).

51. Forrest, A. D. et al. Focal adhesion kinase (FAK): a regulator of CNS myelination. *J. Neurosci. Res.* **87**, 3456–3464 (2009).

52. Itoh, F. et al. Poor vessel formation in embryos from knock-in mice expressing ALK5 with L45 loop mutation defective in Smad activation. *Lab. Invest.* **89**, 800–810 (2009).

Acknowledgements

We acknowledge staff at the Queen Square Brain Bank for Neurological

Disorders, University College London Queen Square Institute of Neurology (Z. Jaunmuktane) for providing ALSP tissue and tissue from unaffected individuals for immunofluorescence analysis; staff at the School of Biological Sciences electron microscopy unit (Wellcome Multi-User Equipment Grant WT104915MA); staff at the Core Facility for Electron Microscopy of the Charité; M. Mohammad for support in data acquisition; and A. Corsinotti and F. Rossi at the CRM Flow Cytometry Core Facility (University of Edinburgh) for their assistance with cell processing for single-cell RNA sequencing. The Queen Square Brain Bank is supported by the Reta Lila Weston Institute of Neurological Studies, University College London Queen Square Institute of Neurology. This work was supported by the Wellcome Trust-funded four-year Ph.D. programme in Tissue Repair at The University of Edinburgh (to N.B.M., 108906/Z/15/Z), a Career Development Award from the Medical Research Council and the United Kingdom Multiple Sclerosis Society (to V.E.M., MR/M020827/1), a Senior Non-Clinical Research Fellowship from the Medical Research Council (to V.E.M., MR/V031260/1), the John David Eaton Chair in Multiple Sclerosis Research at the Barlo Multiple Sclerosis Centre (to V.E.M.), the Medical Research Council Centre for Reproductive Health (MR/N02256/1), the Simons Initiative for the Developing Brain (to C.P.), the United Kingdom Dementia Research Institute (UK DRI) (which receives its funding from UK DRI Ltd, the United Kingdom Medical Research Council, the Alzheimer's Society and Alzheimer's Research United Kingdom) (to J.P., ARUK-PG2016B-6), Alzheimer's Research United Kingdom (to K.H., ARUK-PG2016B-6), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (to T.L.S.-J. number 681181), a Grant-in-Aid of Scientific Research (B) from the Japan Society for the Promotion of Sciences (to R.M., number 22H02962), the Research Foundation of Flanders (FWO Vlaanderen, to J.F.J.B. and J.J.A.H., numbers 1S15519N, G099618FWO and 12J9119N), Interreg V-A EMR (EURLIPIDS, to J.F.J.B. and J.J.A.H., number EMR23), the Belgian Charcot Foundation (to J.F.J.B. and J.J.A.H.), and a MRC/MS Society UK Award (to A.W. and N.B.-C., number MR/T015594/1). Diagrams were created using BioRender.com under a publication licence. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Author information

Author notes

1. These authors contributed equally: Josef Priller, Veronique E. Miron

Authors and Affiliations

1. UK Dementia Research Institute at The University of Edinburgh, Edinburgh, UK

   Niamh B. McNamara, David A. D. Munro, Alana Hoffmann, Rebecca K. Holloway, Irene Molina-Gonzalez, Jamie Rose, Stefan Szymkowiak, Barry W. McColl, Tara L. Spires-Jones, Anna Williams, Josef Priller & Veronique E. Miron

2. Centre for Discovery Brain Sciences, Chancellor’s Building, The University of Edinburgh, Edinburgh, UK

   Niamh B. McNamara, Alana Hoffmann, Rebecca K. Holloway, Irene Molina-Gonzalez, Katharine E. Askew, Jamie Rose, Stefan Szymkowiak, Barry W. McColl, Tara L. Spires-Jones, Karen Horsburgh & Veronique E. Miron

3. Medical Research Council Centre for Reproductive Health, The Queen’s Medical Research Institute, The University of Edinburgh, Edinburgh, UK

   Niamh B. McNamara, Alana Hoffmann, Rebecca K. Holloway, Irene Molina-Gonzalez & Veronique E. Miron

4. Centre for Clinical Brain Sciences, Chancellor’s Building, The University of Edinburgh, Edinburgh, UK

   David A. D. Munro & Josef Priller

5. Centre for Regenerative Medicine, Institute for Regeneration and

Repair, The University of Edinburgh, Edinburgh, UK

Nadine Bestard-Cuche, Jonathan Moss & Anna Williams

6. Departments of Molecular Pharmacology, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Japan

Akiko Uyeda & Rieko Muramatsu

7. Department of Immunology and Infection, Biomedical Research Institute, Hasselt University, Hasselt, Belgium

Jeroen F. J. Bogie & Jerome J. A. Hendriks

8. University MS Centre, Hasselt University, Hasselt, Belgium

Jeroen F. J. Bogie & Jerome J. A. Hendriks

9. Barlo Multiple Sclerosis Centre, St Michael's Hospital, Toronto, Ontario, Canada

Rebecca K. Holloway & Veronique E. Miron

10. Keenan Research Centre for Biomedical Science, St Michael's Hospital, Toronto, Ontario, Canada

Rebecca K. Holloway & Veronique E. Miron

11. Department of Immunology, The University of Toronto, Toronto, Ontario, Canada

Rebecca K. Holloway & Veronique E. Miron

12. Wellcome Trust Centre for Cell Biology, King's Buildings, The University of Edinburgh, Edinburgh, UK

Stephen Mitchell

13. Biological and Veterinary Services, Chancellor's Building, The University of Edinburgh, Edinburgh, UK

William Mungall & Michael Dodds

14. Department of Neuropathology and Neurocure Clinical Research Center, Charité-Universitätsmedizin Berlin, Berlin, Germany

Carsten Dittmayer & Werner Stenzel

15. Institute of Neuropathology, Centre for Basics in NeuroModulation, Faculty of Medicine, University of Freiburg, Freiburg, Germany

Lukas Amann & Marco Prinz

16. Signalling Research Centres BIOSS and CIBSS, University of Freiburg, Freiburg, Germany

Lukas Amann & Marco Prinz

17. Centre for Inflammation Research, The Queen's Medical Research Institute, The University of Edinburgh, Edinburgh, UK

Clare Pridans

18. Simons Initiative for the Developing Brain, Centre for Discovery Brain Sciences, University of Edinburgh, Edinburgh, UK

Clare Pridans

19. Muir Maxwell Epilepsy Centre, University of Edinburgh, Edinburgh, UK

Clare Pridans

20. Department of Psychiatry and Psychotherapy, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany

Josef Priller

21. Neuropsychiatry and Laboratory of Molecular Psychiatry, Charité-Universitätsmedizin Berlin and DZNE, Berlin, Germany

Josef Priller

Contributions

N.B.M. designed the study, performed experiments, analysed and interpreted the data and co-wrote the manuscript. D.A.D.M. processed tissue for single-cell RNA sequencing, performed validation studies and a subset of behavioural analysis. N.B.-C. and A.W. performed bioinformatics analysis of single-cell RNA sequencing data and created the shiny application. A.U. and R.M. generated the *Tgfbr1* conditional knockout mouse model. J.F.J.B. and J.J.A.H. performed the lipidomics analysis. A.H., R.K.H., L.A., S.S., B.W.M. and M.P. assisted in the mouse experiments. I.M.-G. performed analysis on astrocytes. K.E.A. and K.H. supported cognitive task training, data analysis and interpretation. S.M., J.M., A.W., C.D., W.S., J.R. and T.L.S.-J. processed tissue and assisted in electron microscopy. W.M. and M.D. performed perfusions and brain extractions. C.P. generated the *Fire*$^{\Delta/\Delta}$ mice, co-supervised the mouse project and contributed to study direction. A.W. and J.P. assisted in single-cell RNA sequencing interpretation, provided guidance and supervision of the study and participated in manuscript editing. V.E.M. supervised the study, guided experimental design and interpretation and co-wrote the manuscript.

Corresponding author

Correspondence to Veronique E. Miron.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Michelle Monje, Xianhua Piao and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Characterisation of FIRE$^{\Delta/\Delta}$ mouse.

a) Images of perivascular macrophages (CD206+; magenta) in association with blood vessels (CD31+; white)(arrows) in FIRE$^{+/+}$ and FIRE$^{\Delta/\Delta}$ mouse corpus callosum at 1 month of age. Scale bar, 25 µm. **b**) Images of perivascular macrophages (LYVE1+ CD68+) at 3-4 months and 6 months of age in corpus callosum. Scale bar, 75 µm. **c**) Mean density of LYVE1+ CD68+ cells/mm^2 ± s.e.m. at 3-4 months and 6 months of age. Non-significant, *P* > 0.9999 and 0.7698, Mann Whitney and 2-tailed unpaired Student's t-test, respectively. At 3-4 months, n = 5 FIRE$^{+/+}$ and 4 FIRE$^{\Delta/\Delta}$ mice, at 6 months, n = 5 FIRE$^{+/+}$ and 3 FIRE$^{\Delta/\Delta}$ mice. **d**) Flow cytometry gating strategy for assessment of non-microglial myeloid cells (including BAMs; CD11b$^+$ CD45^{hi}) for panels (**e**) and (**f**). **e**) Intensity of expression of CSF1R in FIRE$^{+/+}$ and FIRE$^{\Delta/\Delta}$ mice in CD11b$^+$CD45^{lo} microglia (MG) versus CD11b$^+$CD45^{hi} myeloid cells. **f**) Mean Fluorescence Intensity (MFI) of CSF1R ± s.e.m. in non-microglial myeloid cells (including BAMs) in FIRE$^{+/+}$ and FIRE$^{\Delta/\Delta}$ mice. ****$P < 0.0001$, 2-tailed unpaired Student's t-test. n = 4 FIRE$^{+/+}$ and 5 FIRE$^{\Delta/\Delta}$ mice **g**) Images of astrocytes (SOX9; green;

GFAP+; magenta) at 1 month of age in the corpus callosum. Scale bar, 75 µm. Inset shows magnified view of double positive cells. Scale bar, 25 µm. **h**) Mean SOX9+ GFAP+ cells/mm^2 ± s.e.m. at 1 month of age in the corpus callosum. n = 5 mice per group. Non-significant, *P* = 0.4799, 2-tailed unpaired Student's t-test.

Source data

Extended Data Fig. 2 Myelin protein and axonal assessment in the FIRE$^{\Delta/\Delta}$ mouse.

a) Mean pixel intensity of MAG and MBP ± s.e.m. in corpus callosum at 1 month of age. Non-significant, *P* = 0.6482 and 0.1097 respectively, 2-tailed unpaired Student's t-test. n = 5 mice/group. **b**) Images of CNPase, MOG, and PLP (white) in corpus callosum at 1 month of age. Scale bar, 75 µm. **c**) Mean pixel intensity of CNPase, MOG, and PLP ± s.e.m. in corpus callosum at 1 month of age. Non-significant, *P* = 0.9758, 0.5699 and 0.4551 respectively, 2-tailed unpaired Student's t-test. n = 5 mice/group. **d**) Images of MBP (magenta) and MAG (green) in cerebellum at 1 month of age. Scale bar, 75 µm. **e**) Mean pixel intensity of MBP and MAG ± s.e.m. in cerebellum at 1 month of age. Non-significant, *P* = 0.3067 and 0.5462 respectively, 2-tailed unpaired Student's t-test. n = 5 mice/group. **f**) Images of neurofilament (NF; green) in corpus callosum at 1 month of age. Scale bar, 75 µm. **g**) Mean pixel intensity of NF ± s.e.m. in corpus callosum at 1 month of age. Non-significant, *P* = 0.4623, 2-tailed unpaired Student's t-test. n = 5 FIRE$^{+/+}$ and 6 FIRE$^{\Delta/\Delta}$. **h**) Images of axonal spheroids in corpus callosum of FIRE$^{\Delta/\Delta}$ mice at 3-4 months and 6 months, indicated with arrows. Scale bar, 2 µm and 1 µm, respectively.

Source data

Extended Data Fig. 3 Comparison of myelin across time points.

a) Myelin thickness/axon diameter (µm) in FIRE$^{+/+}$ mice at 1, 3-4, and 6

months of age. ****$P < 0.0001$, Kruskal-Wallis with Dunn's multiple comparisons test. n = 3 mice/group. **b**) Myelin thickness/axon diameter (μm) in FIRE$^{\Delta/\Delta}$ mice at 1, 3-4, and 6 months of age. ****$P < 0.0001$, ns $P = 0.9232$, Kruskal-Wallis with Dunn's multiple comparisons test. 1 month: n = 3 FIRE$^{+/+}$ and 4 FIRE$^{\Delta/\Delta}$; 3-4 months: n = 3 mice/group; 6 months: n = 3 FIRE$^{+/+}$ and 4 FIRE$^{\Delta/\Delta}$. **c**) Myelin thickness/axon diameter (μm) in FIRE$^{+/+}$ and FIRE$^{\Delta/\Delta}$ mice at 1, 3-4, and 6 months of age. Table of *P* values from Kruskal-Wallis with Dunn's multiple comparisons test, 1 month: n = 3 FIRE$^{+/+}$ and 4 FIRE$^{\Delta/\Delta}$ mice; 3-4 months: n = 3 mice/group; 6 months: n = 3 FIRE$^{+/+}$ and 4 FIRE$^{\Delta/\Delta}$. **d**) Inner tongue thickness/axon diameter (μm) in FIRE$^{+/+}$ and FIRE$^{\Delta/\Delta}$ mice at 1, 3-4, and 6 months of age. Table of *P* values from Kruskal-Wallis with Dunn's multiple comparisons test,1 month: n = 3 FIRE$^{+/+}$ and 4 FIRE$^{\Delta/\Delta}$ mice; 3-4 months: n = 3 mice/group; 6 months: n = 3 FIRE$^{+/+}$ and 4 FIRE$^{\Delta/\Delta}$. **e**) Mean myelin thickness (μm)/axon diameter bin ± s.e.m. in FIRE$^{+/+}$ and FIRE$^{\Delta/\Delta}$ mice at 1, 3-4, and 6 months of age.1 month: n = 3 FIRE$^{+/+}$ and 4 FIRE$^{\Delta/\Delta}$ mice; 3-4 months: n = 3 mice/group; 6 months: n = 3 FIRE$^{+/+}$ and 4 FIRE$^{\Delta/\Delta}$. **f**) Mean inner tongue thickness (μm)/axon diameter bin ± s.e.m. in FIRE$^{+/+}$ and FIRE$^{\Delta/\Delta}$ mice at 1, 3-4, and 6 months of age. 1 month: n = 3 FIRE$^{+/+}$ and 4 FIRE$^{\Delta/\Delta}$ mice; 3-4 months: n = 3 mice/group; 6 months: n = 3 FIRE$^{+/+}$ and 4 FIRE$^{\Delta/\Delta}$.

Source data

Extended Data Fig. 4 Learning and memory encoding unimpaired, but cognitive flexibility is compromised, in the FIRE$^{\Delta/\Delta}$ mouse.

a) Training phase of Barnes maze: mice learn to locate Target hole 1 (orange) with underlying escape chamber in spatial learning and memory retrieval trials. Mice tested were 2–4 months of age, with a median age of 119 days old (FIRE$^{+/+}$) and 120 days old (FIRE$^{\Delta/\Delta}$). **b**) Mean primary latency (sec) ± s.e.m. over 6 training days. n = 13 FIRE$^{+/+}$ mice and 10 FIRE$^{\Delta/\Delta}$. Non-significant between genotypes across all training days, Day 1: **$P = 0.0011$; Day 2: $P = 0.9787$; Day 3: $P = 0.9627$; Day 4: $P = 0.9574$; Day 5: $P = 0.9990$;

Day 6: $P > 0.9999$, Repeated measures 2-way ANOVA with Sidak's multiple comparisons test. **c**) Mean primary errors ± s.e.m. during training phase. n = 13 FIRE$^{+/+}$ mice and 10 FIRE$^{\Delta/\Delta}$ mice. Non-significant, Day 1: $P = 0.9649$; Day 2: $P = 0.5968$; Day 3: $P = 0.8884$; Day 4: $P = 0.6028$; Day 5: $P = 0.9215$; Day 6: $P = 0.9437$, Repeated measures 2-way ANOVA with Sidak's multiple comparisons test. **d**) Mean percentage of time spent in the target quadrant± s.e.m., n = 13 FIRE$^{+/+}$ and 10 FIRE$^{\Delta/\Delta}$. Dotted line indicates 25% chance. Non-significant between genotypes, 1hr: $P = 0.7337$; 3d: $P > 0.9999$, 2-way ANOVA with Sidak's multiple comparisons test. **e**) Mean number of nose pokes into holes during 1hr probe test ± s.e.m., n = 13 FIRE$^{+/+}$mice and 9 FIRE$^{\Delta/\Delta}$ mice. Target $^{**}P = 0.0019$, 2-way ANOVA with Sidak's multiple comparisons test. **f**) Mean number of nose pokes into holes during 3d probe test ± s.e.m., n = 13 FIRE$^{+/+}$mice and 9 FIRE$^{\Delta/\Delta}$ mice. Non-significant, $P = 0.9329$, 2-way ANOVA with Sidak's multiple comparisons test. **g**) Reversal phase: Target hole 2 (blue) is 180° from the original target; mice require cognitive flexibility to adapt to the new target. **h**) Mean primary latency (sec) ± s.e.m. over 3 training days. n = 13 FIRE$^{+/+}$ and 9 FIRE$^{\Delta/\Delta}$. Non-significant across all training days between genotypes, Day 1: $P = 0.9999$; Day 2: $P = 0.5186$; Day 3: $P = 0.9622$, Repeated measures 2-way ANOVA with Sidak's multiple comparisons test. **i**) Mean primary errors ± s.e.m. during reversal days. n = 13 FIRE$^{+/+}$ and 9 FIRE$^{\Delta/\Delta}$. Day 1: $^{*}P = 0.0488$, Day 2: $^{*}P = 0.0142$, Day 3: $P = 0.6727$, Repeated measures 2-way ANOVA with Sidak's multiple comparisons test. **j**) Mean percentage of time spent in the target quadrant during reversal probe test ± s.e.m. n = 13 FIRE$^{+/+}$ mice and 9 FIRE$^{\Delta/\Delta}$ mice. Dotted line indicates 25% chance. Non-significant, $P = 0.6933$, 2-tailed unpaired Student's t-test. **k**) Mean number of nose pokes into holes during reversal probe test ± s.e.m. n = 13 FIRE$^{+/+}$ mice and 9 FIRE$^{\Delta/\Delta}$ mice. Non-significant $P = 0.9657$, 2-way ANOVA with Sidak's multiple comparisons test. **l**) Schematic of Open Field test. **m**) Total distance travelled (m) during 10-min Open Field test ± s.e.m. n = 37 mice: n = 15 FIRE$^{+/+}$ mice (7 young, 8 middle-aged); n = 22 FIRE$^{\Delta/\Delta}$ mice (12 young, 10 middle-aged). Young mice = 4–8 weeks of age, Middle aged mice = 11–13 months of age. Non-significant, Young: $P = 0.9997$; Middle-aged: $P > 0.999$, 2-way ANOVA. **n**) Percentage (%) of time spent in the centre of the arena during Open Field test ± s.e.m. n = 37 mice: n = 15 FIRE$^{+/+}$ mice (7 young, 8 middle-aged); n = 22 FIRE$^{\Delta/\Delta}$ mice (12 young, 10 middle-aged). Young mice = 4–8 weeks of age,

Middle aged mice = 11–13 months of age. Non-significant, Young: P = 0.7899; Middle-aged: $P > 0.9995$, 2-way ANOVA. **o**) Mean distance travelled (m) during training days ± s.e.m. n = 13 FIRE$^{+/+}$ mice and 9 FIRE$^{\Delta/\Delta}$ mice. Non-significant, P = 0.9607, >0.9999, >0.9999, 0.9955, 0.6583, 0.6034, Repeated measures 2-way ANOVA with Sidak's multiple comparisons test. **p**) Mean distance travelled (m) during reversal days ± s.e.m. n = 13 FIRE$^{+/+}$ mice and 9 FIRE$^{\Delta/\Delta}$ mice. Non-significant, P = 0.7856, 0.9784, 0.8626, Repeated measures 2-way ANOVA with Sidak's multiple comparisons test. **q**) Mean speed (m/s) during training days ± s.e.m. n = 13 FIRE$^{+/+}$ mice and 9 FIRE$^{\Delta/\Delta}$ mice. Non-significant, P = 0.9998, >0.9999, 0.6667, 0.9995, 0.9831, 0.9995, Repeated measures 2-way ANOVA with Sidak's multiple comparisons test. **r**) Mean speed (m/s) during reversal days ± s.e.m. n = 13 FIRE$^{+/+}$ mice and 9 FIRE$^{\Delta/\Delta}$ mice. Non-significant, P = > 0.9999, 0.9940, 0.1561, Repeated measures 2-way ANOVA with Sidak's multiple comparisons test.

Source data

Extended Data Fig. 5 Post-learning oligodendrogenesis and myelination.

a) Representative images of EdU (magenta) and OLIG2 (white) double positive cells (arrows), with magnified view of dotted outline. Scale bars, 25 µm. **b**) Mean OLIG2+ EdU+ cells/mm^2 ± s.e.m. n = 9 FIRE$^{+/+}$ mice and 4 FIRE$^{\Delta/\Delta}$ mice. Non-significant, P = 0.9696, 2-tailed unpaired Student's t-test. **c**) Mean proportion of OLIG2+ EdU+ cells which are mature oligodendrocytes (CC1+; black) or immature lineage cells (CC1-; grey) ± s.e.m. n = 9 FIRE$^{+/+}$ mice and 4 FIRE$^{\Delta/\Delta}$ mice. Non-significant, CC1+: P > 0.9999; CC1-: P > 0.9999, 1-way ANOVA with Tukey's multiple comparisons test. **d**) Representative images of EdU+ (magenta) CC1+ (green) OLIG2+ (white) triple positive cells (arrows). Scale bar, 25 µm. **e**) Images of corpus callosum 6 weeks post cognitive testing. Scale bar, 1 µm. **f**) Mean number of myelinated axons/mm^2 ± s.e.m. in untrained versus trained mice. n = 3 mice/group in untrained category, 4 trained FIRE$^{+/+}$ mice and 6 trained FIRE$^{\Delta/\Delta}$ mice. FIRE$^{+/+}$ mice *P = 0.0364, FIRE$^{\Delta/\Delta}$ mice non-significant, P =

0.8537, 2-tailed unpaired Student's t-test. **g**) Mean percentage increase in myelinated axons in trained FIRE$^{+/+}$ and FIRE$^{\Delta/\Delta}$ mice. n = 3 mice per group in untrained category, 4 trained FIRE$^{+/+}$ mice and 6 trained FIRE$^{\Delta/\Delta}$ mice. **h**) Correlation between mean myelinated axons per mm^2 and reversal day 1 (RD1) primary errors or average primary errors in FIRE$^{+/+}$ and FIRE$^{\Delta/\Delta}$ mice. Each data point represents an individual mouse. n = 4 FIRE$^{+/+}$ and 6 trained FIRE$^{\Delta/\Delta}$ mice.

Source data

Extended Data Fig. 6 Demyelination in FIRE$^{\Delta/\Delta}$ mice.

a) Images of MAG (green) and PLP (magenta) at 6 months of age. Scale bar, 75 µm. **b**) Mean pixel intensity of MAG ± s.e.m. in corpus callosum at 6 months. Non-significant, *P* = 0.4335, 2-tailed unpaired Student's t-test. n = 5 FIRE$^{+/+}$ and 3 FIRE$^{\Delta/\Delta}$. **c**) Mean pixel intensity of PLP ± s.e.m. in corpus callosum at 6 months. Non-significant, *P* = 0.3471, 2-tailed unpaired Student's t-test. n = 5 FIRE$^{+/+}$ mice and 3 FIRE$^{\Delta/\Delta}$ mice. **d**) Mean inner tongue thickness (µm)/axon diameter bin ± s.e.m. at 6 months of age. n = 3 FIRE$^{+/+}$ and 4 FIRE$^{\Delta/\Delta}$ mice. Non-significant, *P* = 0.4837, 2-tailed unpaired Student's t-test. Mean myelin thickness (µm)/axon diameter bin ± s.e.m. at 6 months. n = 3 FIRE$^{+/+}$ and 4 FIRE$^{\Delta/\Delta}$ mice. >1 µm: **P* = 0.0263, 2-tailed unpaired Student's t-test. **e**) Images of mature oligodendrocytes co-expressing OLIG2 (white) and CC1 (magenta) at 6 months of age. Scale bar, 75 µm. **f**) Mean OLIG2+ CC1+ cells/mm^2 ± s.e.m. n = 5 FIRE$^{+/+}$ and 3 FIRE$^{\Delta/\Delta}$. Non-significant, *P* = 0.6400, 2-tailed unpaired Student's t-test. Mean proportion of cells of the oligodendrocyte lineage (OLIG2+) which are mature (CC1+; black) or immature (CC1-; grey) ± s.e.m. n = 5 FIRE$^{+/+}$ and 3 FIRE$^{\Delta/\Delta}$. Non-significant, CC1+: *P* = 0.9938; CC1-: ns *P* = 0.9938, 1-way ANOVA with Tukey's multiple comparisons test. **g**) Images of mature oligodendrocytes co-expressing OLIG2 (white) and CC1 (magenta) at 3-4 months of age. Scale bar, 75 µm. **h**) Mean OLIG2+ CC1+ cells/mm^2 ± s.e.m. n = 5 mice/group. Non-significant, *P* = 0.9825, 2-tailed unpaired Student's t-test. Mean proportion of cells of the oligodendrocyte lineage (OLIG2+) which are mature (CC1+; black) or immature (CC1-; grey) ± s.e.m. n = 5

mice per group. Non-significant, CC1+: $P = 0.7076$; CC1-: $P = 0.7076$, 1-way ANOVA with Tukey's multiple comparisons test. **i**) Images of $FIRE^{+/+}$ and $FIRE^{\Delta/\Delta}$ mouse corpus callosum at 4.5 months of age indicating onset of demyelination in the latter (magenta asterisks), examples of myelinated axons of medium-large calibre indicated by green asterisks. Scale bars, 5 µm and 1 µm. **j**) Mean number of myelinated axons/mm^2 ± s.e.m. n = 3 mice/group. $^{**}P = 0.0042$, 2-tailed unpaired Student's t-test. **k**) Mean myelin thickness per small (<0.6 µm) or medium-large (>0.6 µm) axon diameter bin ± s.e.m. at 3-4 months of age. n = 3 mice/group. <0.6 µm: non-significant, $P = 0.8978$; >0.6 µm: $^{*}P = 0.0491$, 2-way ANOVA with Sidak's multiple comparisons test. **l**) Mean of diameters (0.7290 µm) of demyelinated axons per representative image ± s.e.m. in $FIRE^{\Delta/\Delta}$ mice. n = 3 mice.

Source data

Extended Data Fig. 7 Microglia depletion in adulthood causes hypermyelination and demyelination.

a) Adult $FIRE^{+/+}$ mice were fed the CSF1R inhibitor PLX5622 in the diet from 2 to 3 months of age. **b**) Representative images of IBA1+ cells (magenta) in the corpus callosum of 3-month old $FIRE^{+/+}$ mice on normal diet versus PLX5622 (PLX) diet from 2 to 3 months of age. Scale bar, 25 µm. **c**) Density of IBA1+ cells/mm^2 ± s.e.m. in $FIRE^{+/+}$ mice on normal diet versus PLX diet at 3 months of age. n = 5 mice on normal diet and 3 mice on PLX diet, $^{***}P = 0.0004$, 2-tailed unpaired Student's t-test, **d**) Representative images of hypermyelination in the corpus callosum following PLX administration from 2 to 3 months of age. Scale bar, 1 µm. **e**) Representative images of enlarged inner tongues (orange) and thicker myelin (asterisks) following PLX administration from 2 to 3 months of age. Scale bar, 1 µm. **f**) Inner tongue thickness (µm) versus axon diameter in $FIRE^{+/+}$ mice on normal diet or PLX diet from 2 to 3 months of age. n = 100 axons/mouse. n = 5 mice on normal diet and 3 mice on PLX diet. $^{***}P < 0.0001$, simple linear regression of intercepts. **g**) Myelin thickness (µm) versus axon diameter in $FIRE^{+/+}$ mice on normal diet or PLX diet from 2 to 3 months of age. n = 100 axons per mouse. n = 5 mice on normal diet and 3 mice on PLX diet. $^{***}P <$

0.0001, simple linear regression of slopes. **h**) Images of mature oligodendrocytes co-expressing OLIG2 (white) and CC1 (magenta) in FIRE$^{+/+}$ mice on normal diet or PLX diet from 2 to 3 months of age. Scale bar, 75 µm. **i**) Mean OLIG2+ CC1+ cells per mm^2 ± s.e.m. in FIRE$^{+/+}$ mice on normal diet or PLX diet from 2 to 3 months of age. Non-significant, *P* = 0.3645, 2-tailed unpaired Student's t-test. Mean proportion of cells of the oligodendrocyte lineage (OLIG2+) which are mature (CC1+; black) or immature (CC1-; grey) ± s.e.m at 3 months. Non-significant, CC1+: *P* = 0.3779; CC1-: *P* = 0.3779, 1-way ANOVA with Tukey's multiple comparisons test. n = 5 mice on normal diet and 3 mice on PLX diet. **j**) Adult FIRE$^{+/+}$ mice were fed the CSF1R inhibitor PLX5622 in the diet from 5 to 6-months of age. **k**) Images of demyelination in the corpus callosum of 6-month-old FIRE$^{+/+}$ mice (asterisks) following PLX administration from 5 to 6-months of age. Scale bar, 1 µm. **l**) Mean number of myelinated axons ± s.e.m. in of FIRE$^{+/+}$ mice on normal diet versus PLX diet from 5 to 6-months of age. n = 3 mice/group. ***P* = 0.0024, 2-tailed unpaired Student's t-test.

Source data

Extended Data Fig. 8 Characterization of ALSP tissue.

a) Images of LYVE1+ (green; top) and IBA1+ (magenta) LYVE1+ macrophages (bottom) in human ALSP and age-matched unaffected control frontal white matter. Scale bar, 50 µm. **b**) Magnified images from dotted outlines in (A) of IBA1+ LYVE1+ perivascular macrophages in human ALSP and unaffected age-matched control frontal white matter. Scale bar, 50 µm. **c**) Mean IBA1+ LYVE1+ cells/mm^2 ± s.e.m. in ALSP and unaffected control. n = 3 cases/group. *P* = 0.1000, Mann Whitney test. **d**) Mean proportion of IBA1+ cells which are perivascular macrophages (LYVE1+; grey) or microglia (LYVE1-; black) ± s.e.m. in ALSP and unaffected control. n = 3 cases/group. ***P* = 0.0034, 1-way ANOVA with Tukey's multiple comparisons test. **e**) Mean number of myelinated axons/axon diameter bin in unaffected control and ALSP patients. n = 2 cases/group. **f**) Mean inner tongue thickness (µm)/axon diameter bin in Control and ALSP patients. n = 2 cases/group. **g**) Mean myelin thickness (µm)/axon diameter bin in unaffected control and ALSP patients. n = 2 cases/group.

Source data

Extended Data Fig. 9 Single-cell RNA sequencing analysis of FIRE$^{\Delta/\Delta}$ mice.

(**a**-**d**) t-SNE (t-distributed stochastic neighbour embedding) plots of pre-filtered dataset showing oligodendrocytes (*Mbp*) versus other cell types: **a**) *Fgfr3* (astrocytes), **b**) *Cd68* (microglia/macrophages), **c**) *Pdgfra* (OPCs) and **d**) *Gpr17* (committed oligodendrocyte precursors), with outline around subsetted oligodendrocytes. (**e**-**h**) t-SNE plots where each point is a cell coloured according to: (**e**) Sum of unique molecular identifiers (UMI) counted, (**f**) Percentage of mitochondrial genes, (**g**) Number of detected genes, (**h**) Filtering: In blue are the cells that were kept for subsequent analysis, and in orange the ones filtered out on the basis of UMIs and percentage of mitochondrial genes, see methods for thresholds. (**i**) t-SNE plot without any batch correction. Each point is a cell that is coloured according to its batch of origin (yellow = 3, purple = 4, orange = 5, red = 6). (**j**) t-SNE plot after batch correction with fastMNN. Each point is a cell that is coloured according to its batch of origin (yellow = 3, purple = 4, orange = 5, red = 6). (**k**) Clustering with four different resolutions, obtained by modifying the number of nearest neighbours (*k*). Each resolution is represented in a separate t-SNE plot. (**l**) Clustree plot created using clustree v0.4.4. Each level in this plot denotes a different resolution used for clustering cells. Lines denote contributions of cells from previous clusters, and the size of the circles represents the number of cells. The final clusters were obtained by merging clusters 2 and 6 and clusters 3 and 5 (pink boxes) from the clustering with k = 100 (light blue).

Extended Data Fig. 10 Influence of TGFβ receptor manipulation on myelin health.

a) Images of TGFβR1 expression (green) by OLIG2+ cells (magenta) in corpus callosum of *Tgfbr1*$^{fl/fl}$ and *Plp*-creERT;*Tgfbr1*$^{fl/fl}$ mice, who received 4-hydroxy tamoxifen (OHT; 100 mg/kg/day, P14–18) and were sacrificed at P28. Scale bar, 20 µm. **b**) Mean relative TGFβR1 intensity ± s.e.m. per

OLIG2+ cell in *Tgfbr1*$^{fl/fl}$ and *Plp*-creER;:*Tgfbr1*$^{fl/fl}$ mice. n = 3 mice/group. *P = 0.011, 2-tailed unpaired Student's t-test. **c**) Mean number of myelinated axons/axon diameter bin ± s.e.m. in wildtype mice (WT; green), *Plp*-creERT;*Tgfbr1*$^{fl/fl}$ mice (magenta) and *Tgfbr1*$^{fl/fl}$ mice (blue). n = 3 mice/group. Non-significant, P = 0.3001, 2-way ANOVA with Sidak's multiple comparisons test. **d**) Mean inner tongue thickness (µm)/axon diameter bin ± s.e.m. in wildtype mice (WT; green), *Plp*-creERT;*Tgfbr1*$^{fl/fl}$ mice (magenta) and *Tgfbr1*$^{fl/fl}$ mice (blue). n-3 mice/group. <0.3 µm WT vs. *Plp*-creERT;*Tgfbr1*$^{fl/fl}$: *P = 0.03950, 2-way ANOVA with Sidak's multiple comparisons test. **e**) Mean myelin tongue thickness (µm)/axon diameter bin ± s.e.m. in wildtype mice (WT; green), *Plp*-creERT;*Tgfbr1*$^{fl/fl}$ mice (magenta) and *Tgfbr1*$^{fl/fl}$ mice (blue). n = 3 mice/group. <0.3 µm WT vs. *Tgfbr1*$^{fl/fl}$: *P = 0.0291; >1 µm WT vs. *Plp*-creERT;*Tgfbr1*$^{fl/fl}$ **P = 0.0037 and *Plp*-creERT;*Tgfbr1*$^{fl/fl}$ vs. *Tgfbr1*$^{fl/fl}$ ***P = 0.0003, 2-way ANOVA with Sidak's multiple comparisons test. **f**) Mean number of myelinated axons ± s.e.m./axon diameter bin in 3-month-old FIRE$^{\Delta/\Delta}$ mice following treatment with vehicle control or SRI-011381 hydrochloride from 2 to 3 months of age. n = 3 Vehicle-treated and 4 SRI-treated mice. Non-significant, P = 0.9202, 2-way ANOVA with Sidak's multiple comparisons test. **g**) Mean inner tongue thickness (µm) ± s.e.m. per axon diameter bin in 3-month-old FIRE$^{\Delta/\Delta}$ mice following treatment with vehicle control or SRI-011381 hydrochloride from 2 to 3 months of age. n = 3 Vehicle-treated and 4 SRI-treated mice. 0.5-1.0 µm: *P = 0.0334, >1 µm: ****P < 0.0001, 2-way ANOVA with Sidak's multiple comparisons test. **h**) Mean myelin thickness (µm) ± s.e.m./axon diameter bin in 3-month-old FIRE$^{\Delta/\Delta}$ mice following treatment with vehicle control or SRI-011381 hydrochloride from 2 to 3 months of age. n = 3 Vehicle-treated and 4 SRI-treated mice. 0.5-1.0 µm: **P = 0.0027, >1 µm: ****P < 0.0001, 2-way ANOVA with Sidak's multiple comparisons test.

Source data

Extended Data Table 1 Clinical information from human tissue analysis

Supplementary information

Reporting Summary

Peer Review File

Supplementary Table 1

Differentially expressed genes enriched in Oligo1 in *Fire*$^{\Delta/\Delta}$ mice. Genes that are enriched in the Oligo1 oligodendrocyte cluster versus other oligodendrocyte clusters, with *P* value (p_val), average $\log_2$(fold change) (avg_Log2FC), percentage of cells in Oligo1 (pct.1) and percentage of cells in other oligodendrocyte clusters (pct.2).

Supplementary Table 2

Top predicted upstream regulators of Oligo1. Top upstream regulators of Oligo1 differentially expressed genes as predicted using ingenuity pathway analysis (IPA), with molecule type, *P* value of overlap and target molecules in dataset.

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 4

Source Data Fig. 5

Source Data Fig. 6

Source Data Extended Data Fig. 1

Source Data Extended Data Fig. 2

Source Data Extended Data Fig. 3

Source Data Extended Data Fig. 4

Source Data Extended Data Fig. 5

Source Data Extended Data Fig. 6

Source Data Extended Data Fig. 7

Source Data Extended Data Fig. 8

Source Data Extended Data Fig. 10

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

McNamara, N.B., Munro, D.A.D., Bestard-Cuche, N. *et al.* Microglia regulate central nervous system myelin growth and integrity. *Nature* **613**, 120–129 (2023). https://doi.org/10.1038/s41586-022-05534-y

- Received: 09 June 2021
- Accepted: 05 November 2022
- Published: 14 December 2022
- Issue Date: 05 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05534-y

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05534-y

- Article
- Published: 04 January 2023

Papers and patents are becoming less disruptive over time

- Michael Park,
- Erin Leahey &
- Russell J. Funk

Nature **volume 613**, pages 138–144 (2023)

- 73k Accesses
- 2683 Altmetric
- Metrics details

Abstract

Theories of scientific and technological change view discovery and invention as endogenous processes[1,2], wherein previous accumulated knowledge enables future progress by allowing researchers to, in Newton's words, 'stand on the shoulders of giants'[3,4,5,6,7]. Recent decades have witnessed exponential growth in the volume of new scientific and technological knowledge, thereby creating conditions that should be ripe for major advances[8,9]. Yet contrary to this view, studies suggest that progress is slowing in several major fields[10,11]. Here, we analyse these claims at scale across six decades, using data on 45 million papers and 3.9 million patents from six large-scale datasets, together with a new quantitative metric—the CD index[12]—that characterizes how papers and patents change networks of citations in science and technology. We find that papers and patents are increasingly less likely to break with the past in ways that push science and technology in new directions. This pattern

holds universally across fields and is robust across multiple different citation- and text-based metrics[1,13,14,15,16,17]. Subsequently, we link this decline in disruptiveness to a narrowing in the use of previous knowledge, allowing us to reconcile the patterns we observe with the 'shoulders of giants' view. We find that the observed declines are unlikely to be driven by changes in the quality of published science, citation practices or field-specific factors. Overall, our results suggest that slowing rates of disruption may reflect a fundamental shift in the nature of science and technology.

Main

Although the past century witnessed an unprecedented expansion of scientific and technological knowledge, there are concerns that innovative activity is slowing[18,19,20]. Studies document declining research productivity in semiconductors, pharmaceuticals and other fields[10,11]. Papers, patents and even grant applications have become less novel relative to prior work and less likely to connect disparate areas of knowledge, both of which are precursors of innovation[21,22]. The gap between the year of discovery and the awarding of a Nobel Prize has also increased[23,24], suggesting that today's contributions do not measure up to the past. These trends have attracted increasing attention from policymakers, as they pose substantial threats to economic growth, human health and wellbeing, and national security, along with global efforts to combat grand challenges such as climate change[25,26].

Numerous explanations for this slowdown have been proposed. Some point to a dearth of 'low-hanging fruit' as the readily available productivity-enhancing innovations have already been made[19,27]. Others emphasize the increasing burden of knowledge; scientists and inventors require ever more training to reach the frontiers of their fields, leaving less time to push those frontiers forward[18,28]. Yet much remains unknown, not merely about the causes of slowing innovative activity, but also the depth and breadth of the phenomenon. The decline is difficult to reconcile with centuries of observation by philosophers of science, who characterize the growth of knowledge as an endogenous process, wherein previous knowledge enables future discovery, a view captured famously in Newton's observation that if he had seen further, it was by 'standing on the shoulders of giants'[3].

Moreover, to date, the evidence pointing to a slowdown is based on studies of particular fields, using disparate and domain-specific metrics[10,11], making it difficult to know whether the changes are happening at similar rates across areas of science and technology. Little is also known about whether the patterns seen in aggregate indicators mask differences in the degree to which individual works push the frontier.

We address these gaps in understanding by analysing 25 million papers (1945–2010) in the Web of Science (WoS) (Methods) and 3.9 million patents (1976–2010) in the United States Patent and Trademark Office's (USPTO) Patents View database (Methods). The WoS data include 390 million citations, 25 million paper titles and 13 million abstracts. The Patents View data include 35 million citations, 3.9 million patent titles and 3.9 million abstracts. Subsequently, we replicate our core findings on four additional datasets—JSTOR, the American Physical Society corpus, Microsoft Academic Graph and PubMed—encompassing 20 million papers. Using these data, we join a new citation-based measure[12] with textual analyses of titles and abstracts to understand whether papers and patents forge new directions over time and across fields.

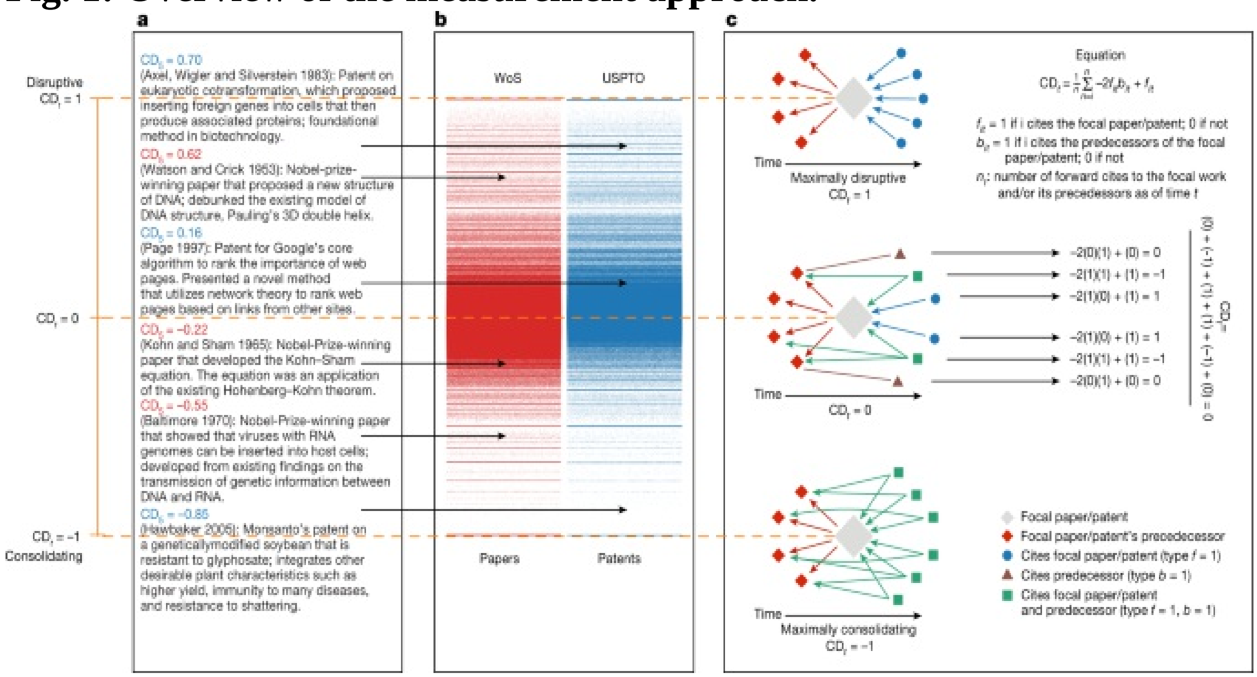
Measurement of disruptiveness

To characterize the nature of innovation, we draw on foundational theories of scientific and technological change[2,29,30], which distinguish between two types of breakthroughs. First, some contributions improve existing streams of knowledge, and therefore consolidate the status quo. Kohn and Sham (1965)[31], a Nobel-winning paper used established theorems to develop a method for calculating the structure of electrons, which cemented the value of previous research. Second, some contributions disrupt existing knowledge, rendering it obsolete, and propelling science and technology in new directions. Watson and Crick (1953)[32], also a Nobel winner, introduced a model of the structure of DNA that superseded previous approaches (for example, Pauling's triple helix). Kohn and Sham and Watson and Crick were both important, but their implications for scientific and technological change were different.

We quantify this distinction using a measure—the CD index[12]—that

characterizes the consolidating or disruptive nature of science and technology (Fig. 1). The intuition is that if a paper or patent is disruptive, the subsequent work that cites it is less likely to also cite its predecessors; for future researchers, the ideas that went into its production are less relevant (for example, Pauling's triple helix). If a paper or patent is consolidating, subsequent work that cites it is also more likely to cite its predecessors; for future researchers, the knowledge upon which the work builds is still (and perhaps more) relevant (for example, the theorems Kohn and Sham used). The CD index ranges from −1 (consolidating) to 1 (disruptive). We measure the CD index five years after the year of each paper's publication (indicated by CD_5, see Extended Data Fig. 1 for the distribution of CD_5 among papers and patents and Extended Data Fig. 2 for analyses using alternative windows)[33]. For example, Watson and Crick and Kohn and Sham both received over a hundred citations within five years of being published. However, the Kohn and Sham paper has a CD_5 of −0.22 (indicating consolidation), whereas the Watson and Crick paper has a CD_5 of 0.62 (indicating disruption). The CD index has been validated extensively in previous research, including through correlation with expert assessments[12,34].

Fig. 1: Overview of the measurement approach.



This figure shows a schematic visualization of the CD index. **a**, CD index

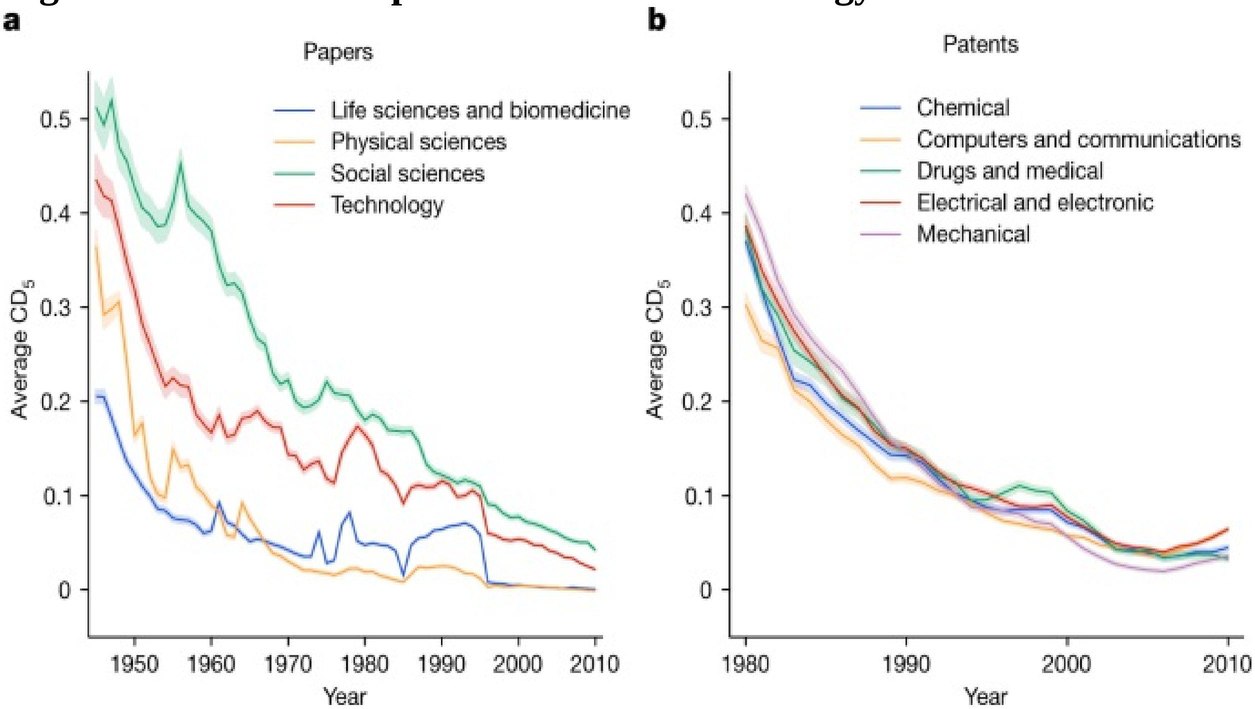
value of three Nobel Prize-winning papers[31,32,58] and three notable patents[59,60,61] in our sample, measured as of five years post-publication (indicated by CD5). **b**, Distribution of CD_5 for papers from WoS (n = 24,659,076) between 1945 and 2010 and patents from Patents View (n = 3,912,353) between 1976 and 2010, where a single dot represents a paper or patent. The vertical (up–down) dimension of each 'strip' corresponds to values of the CD index (with axis values shown in orange on the left). The horizontal (left–right) dimension of each strip helps to minimize overlapping points. Darker areas on each strip plot indicate denser regions of the distribution (that is, more commonly observed CD_5 values). Additional details on the distribution of the CD index are given in Extended Data Fig. 1. **c**, Three hypothetical citation networks, where the CD index is at the maximally disruptive value ($CD_t = 1$), midpoint value ($CD_t = 0$), and maximally consolidating value ($CD_t = -1$). The panel also provides the equation for the CD index and an illustrative calculation.

Declining disruptiveness

Across fields, we find that science and technology are becoming less disruptive. Figure 2 plots the average CD_5 over time for papers (Fig. 2a) and patents (Fig. 2b). For papers, the decrease between 1945 and 2010 ranges from 91.9% (where the average CD_5 dropped from 0.52 in 1945 to 0.04 in 2010 for 'social sciences') to 100% (where the average CD_5 decreased from 0.36 in 1945 to 0 in 2010 for 'physical sciences'); for patents, the decrease between 1980 and 2010 ranges from 78.7% (where the average CD_5 decreased from 0.30 in 1980 to 0.06 in 2010 for 'computers and communications') to 91.5% (where the average CD_5 decreased from 0.38 in 1980 to 0.03 in 2010 for 'drugs and medical'). For both papers and patents, the rates of decline are greatest in the earlier parts of the time series, and for patents, they appear to begin stabilizing between the years 2000 and 2005. For papers, since about 1980, the rate of decline has been more modest in 'life sciences and biomedicine' and physical sciences, and most marked and persistent in social sciences and 'technology'. Overall, however, relative to earlier eras, recent papers and patents do less to push science and technology in new directions. The general similarity in trends we observe across fields is

noteworthy in light of 'low-hanging fruit' theories[19,27], which would probably predict greater heterogeneity in the decline, as it seems unlikely fields would 'consume' their low-hanging fruit at similar rates or times.

Fig. 2: Decline of disruptive science and technology.



a,**b**, Decline in CD_5 over time, separately for papers (**a**, $n = 24{,}659{,}076$) and patents (**b**, $n = 3{,}912{,}353$). For papers, lines correspond to WoS research areas; from 1945 to 2010 the magnitude of decline ranges from 91.9% (social sciences) to 100% (physical sciences). For patents, lines correspond to National Bureau of Economic Research (NBER) technology categories; from 1980 to 2010 the magnitude of decline ranges from 93.5% (computers and communications) to 96.4% (drugs and medical). Shaded bands correspond to 95% confidence intervals. As we elaborate in the Methods, this pattern of decline is robust to adjustment for confounding from changes in publication, citation and authorship practices over time.

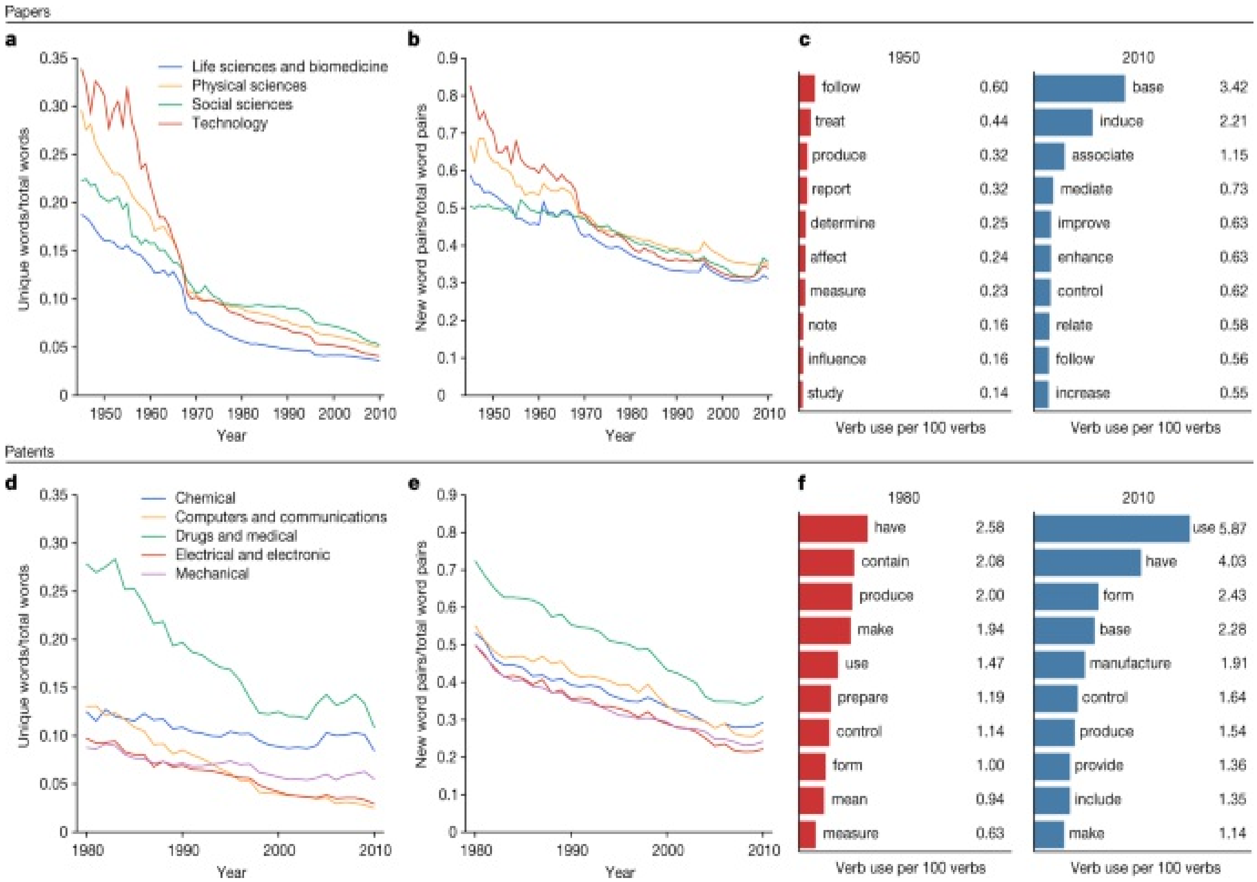
Source data

Linguistic change

The decline in disruptive science and technology is also observable using

alternative indicators. Because they create departures from the status quo, disruptive papers and patents are likely to introduce new words (for example, words used to create a new paradigm might differ from those that are used to develop an existing paradigm)[35,36]. Therefore, if disruptiveness is declining, we would expect a decline in the diversity of words used in science and technology. To evaluate this, Fig. 3a,d documents the type-token ratio (that is, unique/total words) of paper and patent titles over time (Supplementary Information section 1). We observe substantial declines, especially in the earlier periods, before 1970 for papers and 1990 for patents. For paper titles (Fig. 3a), the decrease (1945–2010) ranges from 76.5% (social sciences) to 88% (technology); for patent titles (Fig. 3d), the decrease (1980–2010) ranges from 32.5% (chemical) to 81% (computers and communications). For paper abstracts (Extended Data Fig. 3a), the decrease (1992–2010) ranges from 23.1% (life sciences and biomedicine) to 38.9% (social sciences); for patent abstracts (Extended Data Fig. 3b), the decrease (1980–2010) ranges from 21.5% (mechanical) to 73.2% (computers and communications). In Fig. 3b,e, we demonstrate that these declines in word diversity are accompanied by similar declines in combinatorial novelty; over time, the particular words that scientists and inventors use in the titles of their papers and patents are increasingly likely to have been used together in the titles of previous work. Consistent with these trends in language, we also observe declining novelty in the combinations of previous work cited by papers and patents, based on a previously established measure of ‘atypical combinations’[14] (Extended Data Fig. 4).

Fig. 3: Decline of disruptive science and technology is visible in the changing language of papers and patents.



a,**d**, Figures showing a decline in the diversity of language used in science and technology based on the unique/total words of paper titles from 1945 to 2010 (**a**, *n* = 24,659,076) and of patent titles from 1980 to 2010 (**d**, *n* = 3,912,353). **b**,**e**, Figures showing a decline in the novelty of language used in science and technology based on the number of new word pairs/total word pairs introduced each year in WoS paper titles from 1945 to 2010 (**b**) and in Patents View patent titles from 1980 to 2010 (refs.[1,17]) (**e**). For papers in both **a** and **b**, lines correspond to WoS research areas (*n* = 264 WoS research area × year observations). For patents in both **d** and **e**, lines correspond to NBER technology categories (*n* = 229 NBER technology category × year observations). **c**,**f**, Figures showing the frequency of the most commonly used verbs in paper titles for the first (red) and last (blue) decades of the observation period in paper (**c**, *n* = 24,659,076) and patent (**f**, *n* = 3,912,353) titles.

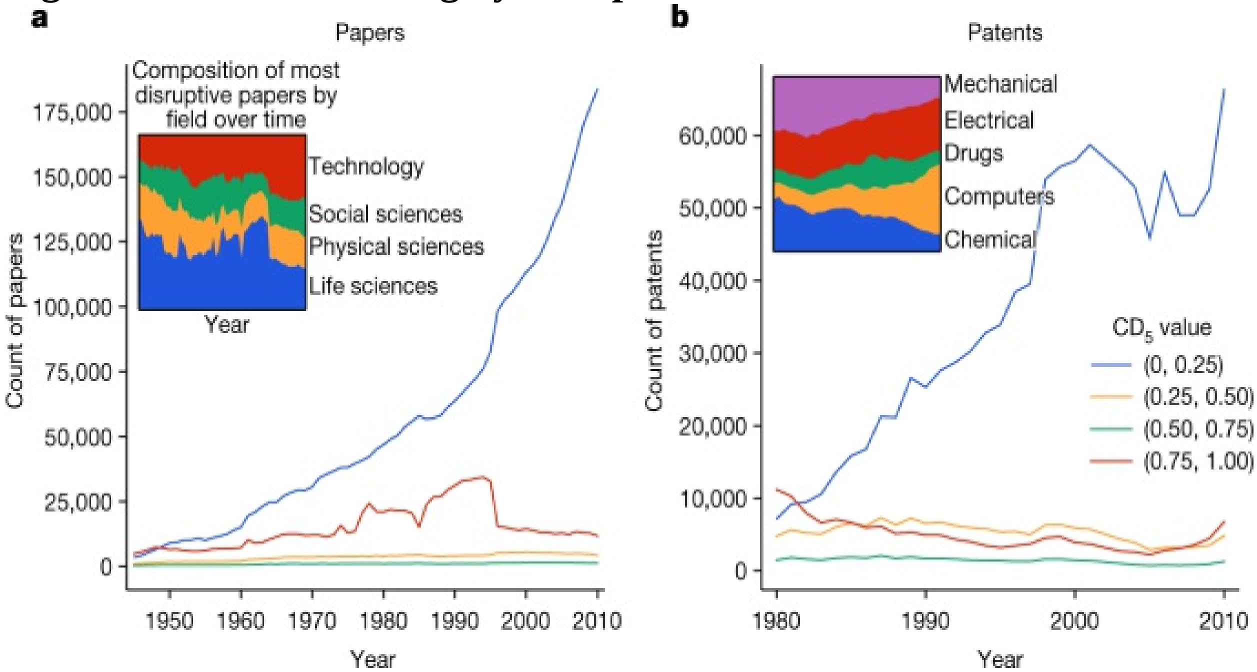
Source data

The decline in disruptive activity is also apparent in the specific words used by scientists and inventors. If disruptiveness is declining, we reasoned that verbs alluding to the creation, discovery or perception of new things should be used less frequently over time, whereas verbs alluding to the improvement, application or assessment of existing things may be used more often[35,36]. Figure 3 shows the most common verbs in paper (Fig. 3c) and patent titles (Fig. 3f) in the first and last decade of each sample (Supplementary Information section 2). Although precisely and quantitatively characterizing words as 'consolidating' or 'disruptive' is challenging in the absence of context, the figure highlights a clear and qualitative shift in language. In the earlier decades, verbs evoking creation (for example, 'produce', 'form', 'prepare' and 'make'), discovery (for example, 'determine' and 'report') and perception (for example, 'measure') are prevalent in both paper and patent titles. In the later decades, however, these verbs are almost completely displaced by those tending to be more evocative of the improvement (for example, 'improve', 'enhance' and 'increase'), application (for example, 'use' and 'include') or assessment (for example, 'associate', 'mediate' and 'relate') of existing scientific and technological knowledge and artefacts. Taken together, these patterns suggest a substantive shift in science and technology over time, with discovery and invention becoming less disruptive in nature, consistent with our results using the CD index.

Conservation of highly disruptive work

The aggregate trends we document mask considerable heterogeneity in the disruptiveness of individual papers and patents and remarkable stability in the absolute number of highly disruptive works (Methods and Fig. 4). Specifically, despite large increases in scientific productivity, the number of papers and patents with CD_5 values in the far right tail of the distribution remains nearly constant over time. This 'conservation' of the absolute number of highly disruptive papers and patents holds despite considerable churn in the underlying fields responsible for producing those works (Extended Data Fig. 5, inset). These results suggest that the persistence of major breakthroughs—for example, measurement of gravity waves and COVID-19 vaccines—is not inconsistent with slowing innovative activity. In short, declining aggregate disruptiveness does not preclude individual highly disruptive works.

Fig. 4: Conservation of highly disruptive work.



This figure shows the number of disruptive papers (**a**, *n* = 5,030,179) and patents (**b**, *n* = 1,476,004) across four different ranges of CD_5 (papers and patents with CD_5 values in the range [−1.0, 0) are not represented in the figure). Lines correspond to different levels of disruptiveness as measured by CD_5. Despite substantial increases in the numbers of papers and patents published each year, there is little change in the number of highly disruptive papers and patents, as evidenced by the relatively flat red, green and orange lines. This pattern helps to account for simultaneous observations of both aggregate evidence of slowing innovative activity and seemingly major breakthroughs in many fields of science and technology. The inset plots show the composition of the most disruptive papers and patents (defined as those with CD_5 values >0.25) by field over time. The observed stability in the absolute number of highly disruptive papers and patents holds despite considerable churn in the underlying fields of science and technology responsible for producing those works. ‘Life sciences’ denotes the life sciences and biomedicine research area; ‘electrical’ denotes the electrical and electronic technology category; ‘drugs’ denotes the drugs and medical technology category; and ‘computers’ denotes the computers and communications technology category.

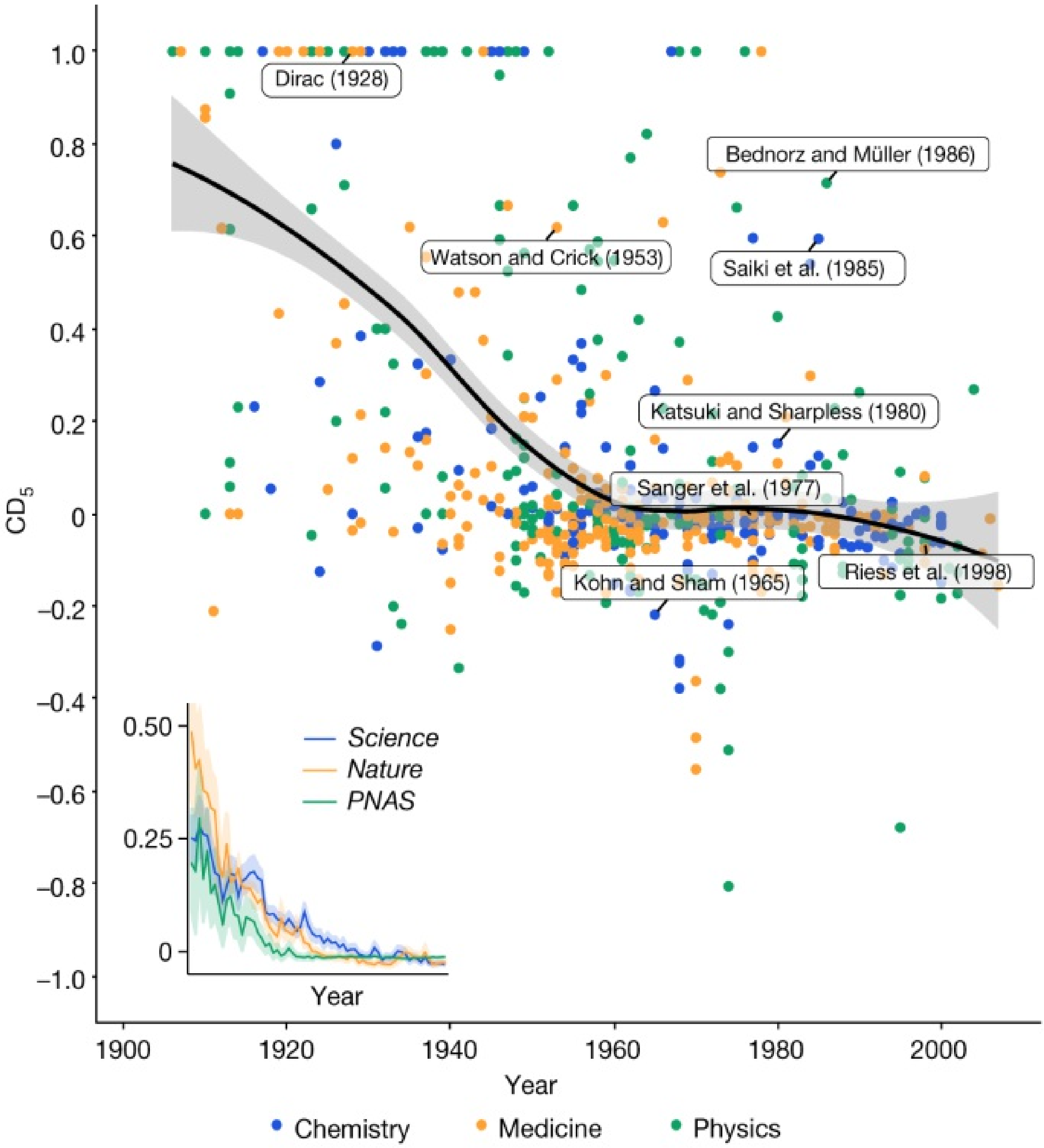
Source data

Alternative explanations

What is driving the decline in disruptiveness? Earlier, we suggested our results are not consistent with explanations that link slowing innovative activity to diminishing ‘low-hanging fruit’. Extended Data Fig. 5 shows that the decline in disruptiveness is unlikely to be due to other field-specific factors by decomposing variation in CD_5 attributable to field, author and year effects (Methods).

Declining rates of disruptive activity are unlikely to be caused by the diminishing quality of science and technology[22,37]. If they were, then the patterns seen in Fig. 2 should be less visible in high-quality work. However, when we restrict our sample to articles published in premier publication venues such as *Nature*, *Proceedings of the National Academy of Sciences* and *Science* or to Nobel-winning discoveries[38] (Fig. 5), the downward trend persists.

Fig. 5: CD index of high-quality science over time.



Dirac (1928): Discovery of the Dirac equation of relativistic quantum theory
Watson and Crick (1953): Discovery of the structure of the DNA
Kohn and Sham (1965): Development of a new method for calculating electronic structure
Sanger et al. (1977): Development of a new method for mapping the order of nucleotides
Katsuki and Sharpless (1980): Development of asymmetric epoxidation
Saiki et al. (1985): Discovery of polymerase chain reaction
Bednorz and Müller (1986): Discovery of superconductivity in cermaic materials
Riess et al. (1998): Discovery of the accelerating expansion of the universe

This figure shows changes in CD_5 over time for papers published in *Nature*, *Proceedings of the National Academy of Sciences* (*PNAS*) and *Science* (inset plot, *n* = 223,745) and Nobel Prize-winning papers (main plot, *n* = 635), with several notable examples[31,32,58,62,63,64,65,66] highlighted. Colours indicate the three different journals in the inset plot; colours indicate the three different fields in which the Nobel Prize is awarded in the main plot. Shaded bands correspond to 95% confidence intervals. For historical completeness, we plot CD index scores for all Nobel papers back to 1900 (the first year in which the prize was awarded); however, our main analyses begin in the post-1945 era, when the WoS data are generally more reliable. The figure indicates that changes in the quality of published science over time is unlikely to be responsible for the decline in disruption.

Source data

Furthermore, the trend is not driven by characteristics of the WoS and UPSTO data or our particular derivation of the CD index; we observe similar declines in disruptiveness when we compute CD_5 on papers in JSTOR, the American Physical Society corpus, Microsoft Academic Graph and PubMed (Methods), the results of which are shown in Extended Data Fig. 6. We further show that the decline is not an artefact of the CD index by reporting similar patterns using alternative derivations[13,15] (Methods and Extended Data Fig. 7).

Declines in disruptiveness are also not attributable to changing publication, citation or authorship practices (Methods). First, using approaches from the bibliometrics literature[39,40,41,42,43], we computed several normalized versions of the CD index that adjusted for the increasing tendency for papers and patents to cite previous work[44,45]. Results using these alternative indicators (Extended Data Fig. 8a,d) were similar to those we reported in Fig. 2. Second, using regression, we estimated models of CD_5 as a function of indicator variables for each paper or patent’s publication year, along with specific controls for field × year level—number of new papers/patents, mean number of papers/patents cited, mean number of authors or inventors per paper—and paper or patent-level—number of papers or patents cited—factors. Predictions from these models indicated a decline in disruptive

papers and patents (Extended Data Fig. 8b,e and Supplementary Table 1) that was consistent with our main results. Finally, using Monte Carlo simulations, we randomly rewired the observed citation networks while preserving key characteristics of scientists' and inventors' citation behaviour, including the number of citations made and received by individual papers and patents and the age gap between citing and cited works. We find that observed CD_5 values are lower than those from the simulated networks (Extended Data Fig. 8c,f), and the gap is widening: over time, papers and patents are increasingly less disruptive than would be expected by chance. Taken together, these additional analyses indicate that the decline in CD_5 is unlikely to be driven by changing publication, citation or authorship practices.

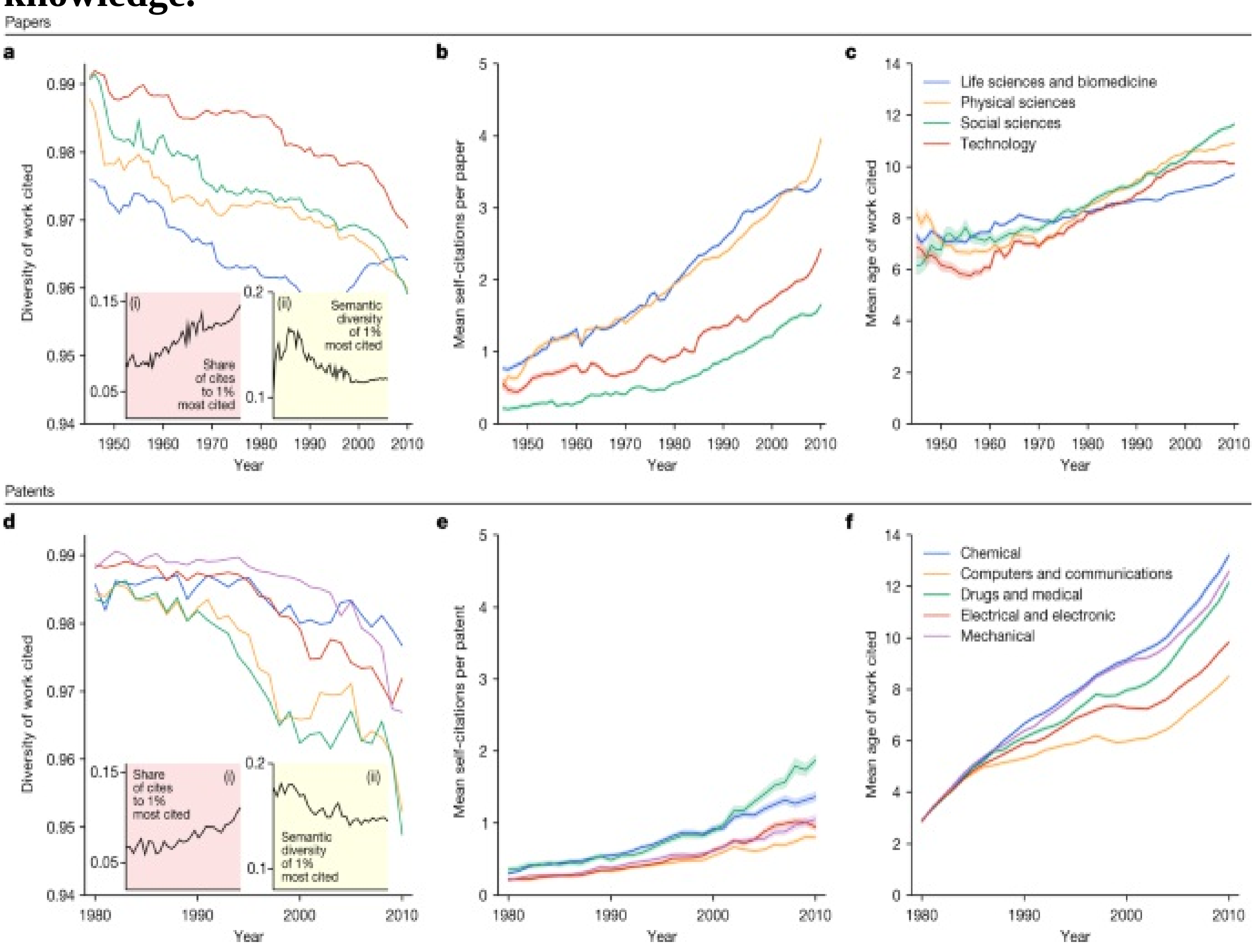
Growth of knowledge and disruptiveness

We also considered how declining disruptiveness relates to the growth of knowledge (Extended Data Fig. 9). On the one hand, scientists and inventors face an increasing knowledge burden, which may inhibit discoveries and inventions that disrupt the status quo. On the other hand, as previously noted, philosophers of science suggest that existing knowledge fosters discovery and invention[3,6,7]. Using regression models, we evaluated the relationship between the stock of papers and patents (a proxy for knowledge) within fields and their CD_5 (Supplementary Information section 3 and Supplementary Table 2). We find a positive effect of the growth of knowledge on disruptiveness for papers, consistent with previous work[20]; however, we find a negative effect for patents.

Given these conflicting results, we considered the possibility that the availability of knowledge may differ from its use. In particular, the growth in publishing and patenting may lead scientists and inventors to focus on narrower slices of previous work[18,46], thereby limiting the 'effective' stock of knowledge. Using three proxies, we document a decline in the use of previous knowledge among scientists and inventors (Fig. 6). First, we see a decline in the diversity of work cited (Fig. 6a,d), indicating that contemporary science and technology are engaging with narrower slices of existing knowledge. Moreover, this decline in diversity is accompanied by an increase in the share of citations to the 1% most highly cited papers and

patents (Fig. 6a (i),d(i)), which are also decreasing in semantic diversity (Fig. 6a (ii),d (ii)). Over time, scientists and inventors are increasingly citing the same previous work, and that previous work is becoming more topically similar. Second, we see an increase in self-citation (Fig. 6b,e), a common proxy for the continuation of one's pre-existing research stream[47,48,49], which is consistent with scientists and inventors relying more on highly familiar knowledge. Third, the mean age of work cited, a common measure for the use of dated knowledge[50,51,52], is increasing (Fig. 6c,f), suggesting that scientists and inventors may be struggling to keep up with the pace of knowledge expansion and instead relying on older, familiar work. All three indicators point to a consistent story: a narrower scope of existing knowledge is informing contemporary discovery and invention.

Fig. 6: Papers and patents are using narrower portions of existing knowledge.



a–**f**, Changes in the level of diversity of existing scientific and technological

knowledge use among papers (**a**, *n* = 264 WoS research area × year observations; **b** and **c**, *n* = 24,659,076 papers) and patents (**d**, 229 NBER technology category × year observations; **e** and **f**, *n* = 3,912,353 patents) based on following measures: diversity of work cited (**a** and **d**), mean number of self-citations (**b** and **e**) and mean age of cited work (**c** and **f**). Shaded bands (**b**,**c**,**e** and **f**) correspond to 95% confidence intervals. The inset plots of **a** and **d** show changes in the share of citations to the top 1% most highly cited papers (**a**(i) and **d**(i)) and in the semantic diversity of the top 1% most cited over time (**a**(ii) and **d**(ii)). Values of both measures are computed within field and year, and are subsequently averaged across fields for plotting. Semantic diversity is based on paper and patent titles; values correspond to the ratio of the standard deviation to the mean pairwise cosine similarity (that is, the coefficient of variation) among the titles of the 1% most cited papers and patents by field and year. To enable semantic comparisons, titles were vectorized using pretrained word embeddings. For papers, lines are shown for each WoS research area; for patents, lines are shown for each NBER technology category. In subsequent regression analyses using these measures, we find that using less diverse work, more of one's own work and older work is associated with less disruptive papers and patents (Methods and Extended Data Table 1).

Source data

Results from a subsequent series of regression models suggest that use of less diverse work, more of one's own work and older work are all negatively associated with disruption (Methods, Extended Data Table 1 and Supplementary Table 3), a pattern that holds even after accounting for the average age and number of previous works produced by team members. When the range of work used by scientists and inventors narrows, disruptive activity declines.

Discussion

In summary, we report a marked decline in disruptive science and technology over time. Our analyses show that this trend is unlikely to be driven by changes in citation practices or the quality of published work. Rather, the decline represents a substantive shift in science and technology, one that

reinforces concerns about slowing innovative activity. We attribute this trend in part to scientists' and inventors' reliance on a narrower set of existing knowledge. Even though philosophers of science may be correct that the growth of knowledge is an endogenous process—wherein accumulated understanding promotes future discovery and invention—engagement with a broad range of extant knowledge is necessary for that process to play out, a requirement that appears more difficult with time. Relying on narrower slices of knowledge benefits individual careers[53], but not scientific progress more generally.

Moreover, even though the prevalence of disruptive works has declined, we find that the sheer number has remained stable. On the one hand, this result may suggest that there is a fixed 'carrying capacity' for highly disruptive science and technology, in which case, policy interventions aimed at increasing such work may prove challenging. On the other hand, our observation of considerable churn in the underlying fields responsible for producing disruptive science and technology suggests the potential importance of factors such as the shifting interests of funders and scientists and the 'ripeness' of scientific and technologicalknowledge for breakthroughs, in which case the production of disruptive work may be responsive to policy levers. In either case, the stability we observe in the sheer number of disruptive papers and patents suggests that science and technology do not appear to have reached the end of the 'endless frontier'. Room remains for the regular rerouting that disruptive works contribute to scientific and technological progress.

Our study is not without limitations. Notably, even though research to date supports the validity of the CD index[12,34], it is a relatively new indicator of innovative activity and will benefit from future work on its behaviour and properties, especially across data sources and contexts. Studies that systematically examine the effect of different citation practices[54,55], which vary across fields, would be particularly informative.

Overall, our results deepen understanding of the evolution of knowledge and may guide career planning and science policy. To promote disruptive science and technology, scholars may be encouraged to read widely and given time to keep up with the rapidly expanding knowledge frontier. Universities may

forgo the focus on quantity, and more strongly reward research quality[56], and perhaps more fully subsidize year-long sabbaticals. Federal agencies may invest in the riskier and longer-term individual awards that support careers and not simply specific projects[57], giving scholars the gift of time needed to step outside the fray, inoculate themselves from the publish or perish culture, and produce truly consequential work. Understanding the decline in disruptive science and technology more fully permits a much-needed rethinking of strategies for organizing the production of science and technology in the future.

Methods

WoS data

We limit our focus to research papers published between 1945 and 2010. Although the WoS data begin in the year 1900, the scale and social organization of science shifted markedly in the post-war era, thereby making comparisons with the present difficult and potentially misleading[67,68,69]. We end our analyses of papers in 2010 because some of our measures require several subsequent years of data following paper publication. The WoS data archive 65 million documents published in 28,968 journals between 1900 and 2017 and 735 million citations among them. In addition, the WoS data include the titles and the full text of abstracts for 65 and 29 million records, respectively, published between 1913 and 2017. After eliminating non-research documents (for example, book reviews and commentaries) and subsetting the data to the 1945–2010 window, the analytical sample consists of $n = 24{,}659{,}076$ papers.

Patents View data

We limit our focus to patents granted from 1976, which is the earliest year for which machine-readable records are available in the Patents View data. As we did with papers, we end our analyses in 2010 because some measures require data from subsequent years for calculation. The Patents View data are the most exhaustive source of historical data on inventions, with information

on 6.5 million patents granted between 1976 and 2017 and their corresponding 92 million citations. The Patents View data include the titles and abstracts for 6.5 million patents granted between 1976 and 2017. Following previous work[12], we focused our attention on utility patents, which cover the vast majority (91% in our data) of patented inventions. After eliminating non-utility patents and subsetting the data to the 1976–2010 window, the analytical sample consists of $n = 3,912,353$ patents.

Highly disruptive papers and patents

Observations (and claims) of slowing progress in science and technology are increasingly common, supported not only by the evidence we report, but also by previous research from diverse methodological and disciplinary perspectives[10,11,18,19,20,21,22,23,24]. Yet as noted in the main text, there is a tension between observations of slowing progress from aggregate data on the one hand, and continuing reports of seemingly major breakthroughs in many fields of science and technology—spanning everything from the measurement of gravity waves to the sequencing of the human genome—on the other. In an effort to reconcile this tension, we considered the possibility that whereas overall, discovery and invention may be less disruptive over time, the high-level view taken in previous work may mask considerable heterogeneity. Put differently, aggregate evidence of slowing progress does not preclude the possibility that some subset of discoveries and inventions is highly disruptive.

To evaluate this possibility, we plot the number of disruptive papers (Fig. 4a) and patents (Fig. 4b) over time, where disruptive papers and patents are defined as those with CD_5 values >0. Within each panel, we plot four lines, corresponding to four evenly spaced intervals—(0, 0.25], (0.25, 0.5], (0.5, 0.75], (0.75, 1.00]—over the positive values of CD_5. The first two intervals therefore correspond to papers and patents that are relatively weakly disruptive, whereas the latter two correspond to those that are more strongly so (for example, where we may expect to see major breakthroughs such as some of those mentioned above). Despite major increases in the numbers of papers and patents published each year, we see little change in the number of highly disruptive papers and patents, as evidenced by the relatively flat red,

green and orange lines. Notably, this ‘conservation’ of disruptive work holds even despite fluctuations over time in the composition of the scientific and technological fields responsible for producing the most disruptive work (Fig. 4, inset plots). Overall, these results help to account for simultaneous observations of both major breakthroughs in many fields of science and technology and aggregate evidence of slowing progress.

Relative contribution of field, year and author or inventor effects

Our results show a steady decline in the disruptiveness of science and technology over time. Moreover, the patterns we observe are generally similar across broad fields of study, which suggests that the factors driving the decline are not unique to specific domains of science and technology. The decline could be driven by other factors, such as the conditions of science and technology at a point in time or the particular individuals who produce science and technology. For example, exogenous factors such as economic conditions may encourage research or invention practices that are less disruptive. Similarly, scientists and inventors of different generations may have different approaches, which may result in greater or lesser tendencies for producing disruptive work. We therefore sought to understand the relative contribution of field, year and author (or inventor) factors to the decline of disruptive science and technology.

To do so, we decomposed the relative contribution of field, year and author fixed effects to the predictive power of regression models of the CD index. The unit of observation in these regressions is the author (or inventor) × year. We enter field fixed effects using granular subfield indicators (that is, 150 WoS subject areas for papers, 138 NBER subcategories for patents). For simplicity, we did not include additional covariates beyond the fixed effects in our models. Field fixed effects capture all field-specific factors that do not vary by author or year (for example, the basic subject matter); year fixed effects capture all year-specific factors that do not vary by field or author (for example, the state of communication technology); author (or inventor) fixed effects capture all author-specific factors that do not vary by field or year (for example, the year of PhD awarding). After specifying our model, we

determine the relative contribution of field, year and author fixed effects to the overall model adjusted R^2 using Shapley–Owen decomposition. Specifically, given our $n = 3$ groups of fixed effects (field, year and author) we evaluate the relative contribution of each set of fixed effects by estimating the adjusted R^2 separately for the 2^n models using subsets of the predictors. The relative contribution of each set of fixed effects is then computed using the Shapley value from game theory[70].

Results of this analysis are shown in Extended Data Fig. 5, for both papers (top bar) and patents (bottom bar). Total bar size corresponds to the value of the adjusted R^2 for the fully specified model (that is, with all three groups of fixed effects). Consistent with our observations from plots of the CD index over time, we observe that for both papers and patents, field-specific factors make the lowest relative contribution to the adjusted R^2 (0.02 and 0.01 for papers and patents, respectively). Author fixed effects, by contrast, appear to contribute much more to the predictive power of the model, for both papers (0.20) and patents (0.17). Researchers and inventors who entered the field in more recent years may face a higher burden of knowledge and thus resort to building on narrower slices of existing work (for example, because of more specialized doctoral training), which would generally lead to less disruptive science and technology being produced in later years, consistent with our findings. The pattern is more complex for year fixed effects; although year-specific factors that do not vary by field or author hold more explanatory power than field for both papers (0.02) and patents (0.16), they appear to be substantially more important for the latter than the former. Taken together, these findings suggest that relatively stable factors that vary across individual scientists and inventors may be particularly important for understanding changes in disruptiveness over time. The results also confirm that domain-specific factors across fields of science and technology play a very small role in explaining the decline in disruptiveness of papers and patents.

Alternative samples

We also considered whether the patterns we document may be artefacts of our choice of data sources. Although we observe consistent trends in both the WoS and Patents View data, and both databases are widely used by the

Science of Science community, our results may conceivably be driven by factors such as changes in coverage (for example, journals added or excluded from WoS over time) or even data errors rather than fundamental changes in science and technology. To evaluate this possibility, we therefore calculated CD_5 for papers in four additional databases—JSTOR, the American Physical Society corpus, Microsoft Academic Graph and PubMed. We included all records from 1930 to 2010 from PubMed (16,774,282 papers), JSTOR (1,703,353 papers) and American Physical Society (478,373 papers). The JSTOR data were obtained via a special request from ITHAKA, the data maintainer (http://www.ithaka.org), as were the American Physical Society data (https://journals.aps.org/datasets). We downloaded the Microsoft Academic Graph data from CADRE at Indiana University (https://cadre.iu.edu/). The PubMed data were downloaded from the National Library of Medicine FTP server (ftp://ftp.ncbi.nlm.nih.gov/pubmed/baseline). Owing to the exceptionally large scale of Microsoft Academic Graph and the associated computational burden, we randomly extracted 1 million papers. As shown in Extended Data Fig. 6, the downward trend in disruptiveness is evident across all samples.

Alternative bibliometric measures

Several recent papers have introduced alternative specifications of the CD index[12]. We evaluated whether the declines in disruptiveness we observe are corroborated using two alternative variations. One criticism of the CD index has been that the number of papers that cite only the focal paper's references dominates the measure[13]. Bornmann et al.[13] proposes $\mathrm{DI}_{l}^{\mathrm{nok}}$ as a variant that is less susceptible to this issue. Another potential weakness of the CD index is that it could be very sensitive to small changes in the forward citation patterns of papers that make no backward citations[15]. Leydesdorff et al.[15] suggests DI* as an alternate indicator of disruption that addresses this issue. Therefore, we calculated $\mathrm{DI}_{l}^{\mathrm{nok}}$ where $l = 5$ and DI* for 100,000 randomly drawn papers and patents each from our analytic sample. Results are presented in Extended Data Fig. 7a (papers) and b (patents). The blue lines indicate disruption based on Bornmann et al.[13] and the orange lines indicate disruption based on Leydesdorff et al.[15]. Across science and

technology, the two alternative measures both show declines in disruption over time, similar to the patterns observed with the CD index. Taken together, these results suggest that the declines in disruption we document are not an artefact of our particular operationalization.

Robustness to changes in publication, citation and authorship practices

We also considered whether our results may be attributable to changes in publication, citation or authorship practices, rather than by substantive shifts in discovery and invention. Perhaps most critically, as noted in the main text, there has been a marked expansion in publishing and patenting over the period of our study. This expansion has naturally increased the amount of previous work that is relevant to current science and technology and therefore at risk of being cited, a pattern reflected in the marked increase in the average number of citations made by papers and patents (that is, papers and patents are citing more previous work than in previous eras)[44,45]. Recall that the CD index quantifies the degree to which future work cites a focal work together with its predecessors (that is, the references in the bibliography of the focal work). Greater citation of a focal work independently of its predecessors is taken to be evidence of a social process of disruption. As papers and patents cite more previous work, however, the probability of a focal work being cited independently of its predecessors may decline mechanically; the more citations a focal work makes, the more likely future work is to cite it together with one of its predecessors, even by chance. Consequently, increases in the number of papers and patents available for citing and in the average number of citations made by scientists and inventors may contribute to the declining values of the CD index. In short, given the marked changes in science and technology over our long study window, the CD index of papers and patents published in earlier periods may not be directly comparable to those of more recent vintage, which could in turn render our conclusions about the decline in disruptive science and technology suspect. We addressed these concerns using three distinctive but complementary approaches—normalization, regression adjustment and simulation.

Verification using normalization

First, following common practice in bibliometric research[39,40,41,42,43], we developed two normalized versions of the CD index, with the goal of facilitating comparisons across time. Among the various components of the CD index, we focused our attention on the count of papers or patents that only cite the focal work's references ($N_{\rm k}$), as this term would seem most likely to scale with the increases in publishing and patenting and in the average number of citations made by papers and patents to previous work[13]. Larger values of $N_{\rm k}$ lead to smaller values of the CD index. Consequently, marked increases in $N_{\rm k}$ over time, particularly relative to other components of the measure, may lead to a downward bias, thereby inhibiting our ability to accurately compare disruptive science and technology in later years with earlier periods.

Our two normalized versions of the CD index aim to address this potential bias by attenuating the effect of increases in $N_{\rm k}$. In the first version, which we call 'Paper normalized', we subtract from $N_{\rm k}$ the number of citations made by the focal paper or patent to previous work ($N_{\rm b}$). The intuition behind this adjustment is that when a focal paper or patent cites more previous work, $N_{\rm k}$ is likely to be larger because there are more opportunities for future work to cite the focal paper or patent's predecessors. This increase in $N_{\rm k}$ would result in lower values of the CD index, although not necessarily as a result of the focal paper or patent being less disruptive. In the second version, which we call 'field × year normalized', we subtract $N_{\rm k}$ by the average number of backward citations made by papers or patents in the focal paper or patent's WoS research area or NBER technology category, respectively, during its year of publication (we label this quantity \({N}_{{\rm{b}}}^{{\rm{m}}{\rm{e}}{\rm{a}}{\rm{n}}}\)). The intuition behind this adjustment is that in fields and time periods in which there is a greater tendency for scientists and inventors to cite previous work, $N_{\rm k}$ is also likely to be larger, thereby leading to lower values of the CD index, although again not necessarily as a result of the focal paper or patent being less disruptive. In cases in which either $N_{\rm b}$ or \({N}_{{\rm{b}}}^{{\rm{m}}{\rm{e}}{\rm{a}}{\rm{n}}}\) exceed the value of $N_{\rm k}$, we set $N_{\rm k}$ to 0 (that is, $N_{\rm k}$ is never negative in the normalized measures). Both adaptations of the CD index are inspired by established approaches in the scientometrics literature, and may be

understood as a form of ‘citing side normalization’ (that is, normalization by correcting for the effect of differences in lengths of references lists)[40].

In Extended Data Fig. 8, we plot the average values of both normalized versions of the CD index over time, separately for papers (Extended Data Fig. 8a) and patents (Extended Data Fig. 8d). Consistent with our findings reported in the main text, we continue to observe a decline in the CD index over time, suggesting that the patterns we observe in disruptive science and technology are unlikely to be driven by changes in citation practices.

Verification using regression adjustment

Second, we adjusted for potential confounding using a regression-based approach. This approach complements the bibliometric normalizations just described by allowing us to account for a broader array of changes in publication, citation and authorship practices in general (the latter of which is not directly accounted for in either the normalization approach or the simulation approach described next), and increases the amount of previous work that is relevant to current science and technology in particular. In Supplementary Table 1, we report the results of regression models predicting CD_5 for papers (Models 1–4) and patents (Models 5–8), with indicator variables included for each year of our study window (the reference categories are 1945 and 1980 for papers and patents, respectively). Models 1 and 4 are the baseline models, and include no other adjustments beyond the year indicators. In Models 2 and 5, we add subfield fixed effects (WoS subject areas for papers and NBER technology subcategories for patents). Finally, in Models 3–4 and 7–8, we add control variables for several field × year level—number of new papers orpatents, mean number of papers or patents cited, mean number of authors or inventors per paper—and paper- or patent-level—number of papers or patents cited—characteristics, thereby enabling more robust comparisons in patterns of disruptive science and technology over the long time period spanned by our study. For the paper models, we also include a paper-level control for the number of unlinked references (that is, the number of citations to works that are not indexed in WoS). We find that the inclusion of these controls improves model fit, as indicated by statistically significant Wald tests presented below the relevant models.

Across all eight models shown in Supplementary Table 1, we find that the coefficients on the year indicators are statistically significant and negative, and growing in magnitude over time, which is consistent with the patterns we reported based on unadjusted CD_5 values index in the main text (Fig. 2). In Extended Data Fig. 8, we visualize the results of our regression-based approach by plotting the predicted CD_5 values separately for each of the year indicators included in Models 4 (papers) and 8 (patents). To enable comparisons with raw CD_5 values shown in the main text, we present the separate predictions made for each year as a line graph. As shown in the figure, we continue to observe declining values of the CD index across papers and patents, even when accounting for changes in publication, citation and authorship practices.

Verification using simulation

Third, following related work in the Science of Science[14,71,72,73], we considered whether our results may be an artefact of changing patterns in publishing and citation practices by using a simulation approach. In essence, the CD index measures disruption by characterizing the network of citations around a focal paper or patent. However, many complex networks, even those resulting from random processes, exhibit structures that yield non-trivial values on common network measures (for example, clustering)[74,75,76]. During the period spanned by our study, the citation networks of science and technology experienced significant change, with marked increases in both the numbers of nodes (that is, papers or patents) and edges (that is, citations). Thus, rather than reflecting a meaningful social process, the observed declines in disruption may result from these structural changes in the underlying citation networks.

To evaluate this possibility, we followed standard techniques from network science[75,77] and conducted an analysis in which we recomputed the CD index on randomly rewired citation networks. If the patterns we observe in the CD index are the result of structural changes in the citation networks of science and technology (for example, growth in the number of nodes or edges) rather than a meaningful social process, then these patterns should also be visible in comparable random networks that experience similar structural changes.

Therefore, finding that the patterns we see in the CD index differ for the observed and random citation networks would serve as evidence that the decline in disruption is not an artefact of the data.

We began by creating copies of the underlying citation network on which the values of the CD index used in all analyses reported in the main text were based, separately for papers and patents. For each citation network (one for papers, one for patents), we then rewired citations using a degree-preserving randomization algorithm. In each iteration of the algorithm, two edges (for example, A–B and C–D) are selected from the underlying citation network, after which the algorithm attempts to swap the two endpoints of the edges (for example, A–B becomes A–D, and C–D becomes C–B). If the degree centrality of A, B, C and D remains the same after the swap, the swap is retained; otherwise, the algorithm discards the swap and moves on to the next iteration. When evaluating degree centrality, we consider 'in-degree' (that is, citations from other papers or patents to the focal paper or patent) and 'out-degree' (that is, citations from the focal paper or patent to other papers or patents) separately. Furthermore, we also required that the age distribution of citing and cited papers or patents was identical in the original and rewired networks. Specifically, swaps were only retained when the publication year of the original and candidate citations was the same. In light of these design choices, our rewiring algorithm should be seen as fairly conservative, as it preserves substantial structure from the original network. There is no scholarly consensus on the number of swaps necessary to ensure the original and rewired networks are sufficiently different from one another; the rule we adopt here is $100 \times m$, where m is the number of edges in the network being rewired.

Following previous work[14], we created ten rewired copies of the observed citation networks for both papers and patents. After creating these rewired citation networks, we then recomputed CD_5. Owing to the large scale of the WoS data, we base our analyses on a random subsample of ten million papers; CD_5 was computed on the rewired network for all patents. For each paper and patent, we then compute a z score that compares the observed CD_5 value to those of the same paper or patent in the ten rewired citation networks. Positive z scores indicate that the observed CD_5 value is greater (that is, more disruptive) than would be expected by chance; negative z scores

indicate that the observed values are lesser (that is, more consolidating).

The results of these analyses are shown in Extended Data Fig. 8, separately for papers (Extended Data Fig. 8c) and patents (Extended Data Fig. 8f). Lines correspond to the average *z* score among papers or patents published in the focal year. The plots reveal a pattern of change in the CD index over and beyond that 'baked in' to the changing structure of the network. We find that on average, papers and patents tend to be less disruptive than would be expected by chance, and moreover, the gap between the observed CD index values and those from the randomly rewired networks is increasing over time, which is consistent with our findings of a decline in disruptive science and technology.

Taken together, the results of the foregoing analyses suggest that although there have been marked changes in science and technology over the course of our long study window, particularly with respect to publication, citation and authorship practices, the decline in disruptive science and technology that we document using the CD index is unlikely to be an artefact of these changes, and instead represents a substantive shift in the nature of discovery and invention.

Regression analysis

We evaluate the relationship between disruptiveness and the use of previous knowledge using regression models, predicting CD_5 for individual papers and patents, based on three indicators of previous knowledge use—the diversity of work cited, mean number of self-citations and mean age of work cited. Our measure of the diversity of work cited is measured at the field × year level; all other variables included in the regressions are defined at the level of the paper or patent. To account for potential confounding factors, our models included year and field fixed effects. Year fixed effects account for time variant factors that affect all observations (papers or patents) equally (for example, global economic trends). Field fixed effects account for field-specific factors that do not change over time (for example, some fields may intrinsically value disruptive work over consolidating ones). In contrast to our descriptive plots, for our regression models, we adjust for field effects using the more granular 150 WoS 'extended subjects' (for example, 'biochemistry

and molecular biology’, ‘biophysics’, ‘biotechnology and applied microbiology’, ‘cell biology’, ‘developmental biology’, ‘evolutionary biology’ and ‘microbiology’ are extended subjects within the life sciences and biomedicine research area) and 38 NBER technology subcategories (for example, ‘agriculture’, ‘food’, ‘textile’; ‘coating’; ‘gas’; ‘organic’; and ‘resins’ are subcategories within the chemistry technology category).

In addition, we also include controls for the ‘mean age of team members’ (that is, ‘career age’, defined as the difference between the publication year of the focal paper or patent and the first year in which each author or inventor published a paper or patent) and the ‘mean number of previous works produced by team members’. Although increases in rates of self-citations may indicate that scientists and inventors are becoming more narrowly focused on their own work, these rates may also be driven in part by the amount of previous work available for self-citing. Similarly, although increases in the age of work cited in papers and patents may indicate that scientists and inventors are struggling to keep up, they may also be driven by the rapidly aging workforce in science and technology[78,79]. For example, older scientists and inventors may be more familiar with or more attentive to older work, or may actively resist change[80]. These control variables help to account for these alternative explanations.

Supplementary Table 3 shows summary statistics for variables used in the ordinary-least-squares regression models. The diversity of work cited is measured by normalized entropy, which ranges from 0 to 1. Greater values on this measure indicate a more uniform distribution of citations to a wider range of existing work; lower values indicate a more concentrated distribution of citations to a smaller range of existing work. The tables show that the normalized entropy in a given field and year has a nearly maximal average entropy of 0.98 for both science and technology. About 16% of papers cited in a paper are by an author of the focal paper; the corresponding number for patents is about 7%. Papers tend to rely on older work and work that varies more greatly in age (measured by standard deviation) than patents. In addition, the average CD_5 of a paper is 0.04 whereas the average CD_5 of a patent is 0.12, meaning that the average paper tends to be less disruptive than the average patent.

We find that using more diverse work, less of one's own work and older work tends to be associated with the production of more disruptive science and technology, even after accounting for the average age and number of previous works produced by team members. These findings are based on our regression results, shown in Extended Data Table 1. Models 6 and 12 present the full regression models. The models indicate a consistent pattern for both science and technology, wherein the coefficients for diversity of work cited are positive and significant for papers (0.159, $P < 0.01$) and patents (0.069, $P < 0.01$), indicating that in fields in which there is more use of diverse work, there is greater disruption. Holding all other variables at their means, the predicted CD_5 of papers and patents increases by 303.5% and 1.3%, respectively, when the diversity of work cited increases by 1 s.d. The coefficients of the ratio of self-citations to total work cited is negative and significant for papers (−0.011, $P < 0.01$) and patents (−0.060, $P < 0.01$), showing that when researchers or inventors rely more on their own work, discovery and invention tends to be less disruptive. Again holding all other variables at their means, the predicted CD_5 of papers and patents decreases by 622.9% and 18.5%, respectively, with a 1 s.d. increase in the ratio. The coefficients of the interaction between mean age of work cited and dispersion in age of work cited is positive and significant for papers (0.000, $P < 0.01$) and patents (0.001, $P < 0.01$), suggesting that—holding the dispersion of the age of work cited constant—papers and patents that engage with older work are more likely to be disruptive. The predicted CD_5 of papers and patents increases by a striking 2,072.4% and 58.4%, respectively, when the mean age of work cited increases by 1 s.d. (about nine and eight years for papers and patents, respectively), again holding all other variables at their means. In summary, the regression results suggest that changes in the use of previous knowledge may contribute to the production of less disruptive science and technology.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data associated with this study are freely available in a public repository at https://doi.org/10.5281/zenodo.7258379. Our study draws on data from six sources: the American Physical Society, JSTOR, Microsoft Academic Graph, Patents View, PubMed and WoS. Data from Microsoft Academic Graph, Patents View and PubMed are publicly available, and our repository includes complete data for analyses from these sources. Data from the American Physical Society, JSTOR and WoS are not publicly available, and were used under licence from their respective publishers. To facilitate replication, our repository includes limited versions of the data from these sources, which will enable calculation of basic descriptive statistics. The authors will make full versions of these data available upon request and with permission from their respective publishers. Source data are provided with this paper.

Code availability

Open-source code related to this study is available at https://doi.org/10.5281/zenodo.7258379 and http://www.cdindex.info. We used Python v.3.10.6 (pandas v.1.4.3, numpy v.1.23.1, matplotlib v.3.5.2, seaborn v.0.11.2, spacy v.2.2, jupyterlab v.3.4.4) to wrangle, analyse and visualize data and to conduct statistical analyses. We used MariaDB v.10.6.4 to wrangle data. We used R v.4.2.1 (ggplot2 v.3.36, ggrepel v.0.9.0) to visualize data. We used StataMP v.17.0 (reghdfe v.5.7.3) to conduct statistical analyses.

References

1. Fleming, L. Recombinant uncertainty in technological search. *Manage. Sci*. **47**, 117–132 (2001).

2. Schumpeter, J. *Capitalism, Socialism and Democracy* (Perennial, 1942).

3. Koyré, A. An unpublished letter of Robert Hooke to Isaac Newton. *ISIS* **43**, 312–337 (1952).

4. Popper, K. *Conjectures and Refutations: The Growth of Scientific Knowledge* (Routledge, 2014).

5. Fleck, L. *Genesis and Development of a Scientific Fact* (Univ. Chicago Press, 2012).

6. Acemoglu, D., Akcigit, U. & Kerr, W. R. Innovation network. *Proc. Natl Acad. Sci. USA* **113**, 11483–11488 (2016).

7. Weitzman, M. L. Recombinant growth. *Q. J. Econ.* **113**, 331–360 (1998).

8. Tria, F., Loreto, V., Servedio, V. D. P. & Strogatz, S. H. The dynamics of correlated novelties. *Sci. Rep.* **4**, 1–8 (2014).

9. Fink, T. M. A., Reeves, M., Palma, R. & Farr, R. S. Serendipity and strategy in rapid innovation. *Nat. Commun.* **8**, 1–9 (2017).

10. Pammolli, F., Magazzini, L. & Riccaboni, M. The productivity crisis in pharmaceutical R&D. *Nat. Rev. Drug Discov.* **10**, 428–438 (2011).

11. Bloom, N., Jones, C. I., Van Reenen, J. & Webb, M. Are ideas getting harder to find? *Am. Econ. Rev.* **110**, 1104–1144 (2020).

12. Funk, R. J. & Owen-Smith, J. A dynamic network measure of technological change. *Manage. Sci.* **63**, 791–817 (2017).

13. Bornmann, L., Devarakonda, S., Tekles, A. & Chacko, G. Are disruption index indicators convergently valid? The comparison of several indicator variants with assessments by peers. *Quant. Sci. Stud.* **1**, 1242–1259 (2020).

14. Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).

15. Leydesdorff, L., Tekles, A. & Bornmann, L. A proposal to revise the disruption index. *Prof. Inf.* **30**, e300121 (2021).

16. Lu, C. et al. Analyzing linguistic complexity and scientific impact. *J. Informetr.* **13**, 817–829 (2019).

17. Hofstra, B. et al. The diversity–innovation paradox in science. *Proc.*

Natl Acad. Sci. USA **117**, 9284–9291 (2020).

18. Jones, B. F. The burden of knowledge and the ‘death of the renaissance man’: is innovation getting harder? *Rev. Econ. Stud.* **76**, 283–317 (2009).

19. Gordon, R. J. *The Rise and Fall of American Growth* (Princeton Univ. Press, 2016).

20. Chu, J. S. G. & Evans, J. A. Slowed canonical progress in large fields of science. *Proc. Natl Acad. Sci. USA* **118**, e2021636118 (2021).

21. Packalen, M. & Bhattacharya, J. NIH funding and the pursuit of edge science. *Proc. Natl Acad. Sci. USA* **117**, 12011–12016 (2020).

22. Jaffe, A. B. & Lerner, J. *Innovation and its Discontents: How Our Broken Patent System Is Endangering Innovation and Progress, and What To Do About It* (Princeton Univ. Press, 2011).

23. Horgan, J. *The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age* (Basic Books, 2015).

24. Collison, P. & Nielsen, M. *Science Is Getting Less Bang for its Buck* (Atlantic, 2018).

25. Nolan, A. Artificial intelligence and the future of science. *oecd.ai*, https://oecd.ai/en/wonk/ai-future-of-science (25 October 2021).

26. *Effective Policies to Foster High-risk/High-reward Research.* OECD Science, Technology, and Industry Policy Papers (OECD, 2021).

27. Cowen, T. *The Great Stagnation: How America Ate All the Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better* (Penguin, 2011).

28. Einstein, A. *The World As I See It* (Citadel Press, 1949).

29. Arthur, W. B. The structure of invention. *Res. Policy* **36**, 274–287 (2007).

30. Tushman, M. L. & Anderson, P. Technological discontinuities and organizational environments. *Adm. Sci. Q.* **31**, 439–465 (1986).

31. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133 (1965).

32. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).

33. Bornmann, L. & Tekles, A. Disruption index depends on length of citation window. *Prof. Inf.* **28**, e280207 (2019).

34. Wu, L., Wang, D. & Evans, J. A. Large teams develop and small teams disrupt science and technology. *Nature* **566**, 378–382 (2019).

35. Kuhn, T. S. *The Structure of Scientific Revolutions* (Univ. Chicago Press, 1962).

36. Brad Wray, K. Kuhn and the discovery of paradigms. *Philos. Soc. Sci.* **41**, 380–397 (2011).

37. Ioannidis, J. P. A. Why most published research findings are false. *PLoS Med.* **2**, e124 (2005).

38. Li, J., Yin, Y., Fortunato, S. & Wang, D. A dataset of publication records for Nobel laureates. *Sci. Data* **6**, 1–10 (2019).

39. Bornmann, L. & Marx, W. Methods for the generation of normalized citation impact scores in bibliometrics: which method best reflects the judgements of experts? *J. Informetr.* **9**, 408–418 (2015).

40. Waltman, L. A review of the literature on citation impact indicators. *J. Informetr.* **10**, 365–391 (2016).

41. Waltman, L. & van Eck, N. J. in *Springer Handbook of Science and Technology Indicators* (eds. Glänzel, W. et al.) 281–300 (Springer, 2019).

42. Bornmann, L. How can citation impact in bibliometrics be normalized?

A new approach combining citing-side normalization and citation percentiles. *Quant. Sci. Stud.* **1**, 1553–1569 (2020).

43. Petersen, A. M., Pan, R. K., Pammolli, F. & Fortunato, S. Methods to account for citation inflation in research evaluation. *Res. Policy* **48**, 1855–1865 (2019).

44. Bornmann, L. & Mutz, R. Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. *J. Assoc. Inf. Sci. Technol.* **66**, 2215–2222 (2015).

45. Bornmann, L., Haunschild, R. & Mutz, R. Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. *Humanit. Soc. Sci. Commun.* **8**, 1–15 (2021).

46. Jones, B. F. & Weinberg, B. A. Age dynamics in scientific creativity. *Proc. Natl Acad. Sci. USA* **108**, 18910–18914 (2011).

47. Bonzi, S. & Snyder, H. Motivations for citation: a comparison of self citation and citation to others. *Scientometrics* **21**, 245–254 (1991).

48. Fowler, J. & Aksnes, D. Does self-citation pay? *Scientometrics* **72**, 427–437 (2007).

49. King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J. & West, J. D. Men set their own cites high: gender and self-citation across fields and over time. *Socius* **3**, 2378023117738903 (2017).

50. Mukherjee, S., Romero, D. M., Jones, B. & Uzzi, B. The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: the hotspot. *Sci. Adv.* **3**, e1601315 (2017).

51. Merton, R. K. Singletons and multiples in scientific discovery: a chapter in the sociology of science. *Proc. Am. Philos. Soc.* **105**, 470–486 (1961).

52. Wang, D., Song, C. & Barabási, A.-L. Quantifying long-term scientific

impact. *Science* **342**, 127–132 (2013).

53. Leahey, E. Not by productivity alone: how visibility and specialization contribute to academic earnings. *Am. Sociol. Rev.* **72**, 533–561 (2007).

54. Tahamtan, I. & Bornmann, L. Core elements in the process of citing publications: conceptual overview of the literature. *J. Informetr.* **12**, 203–216 (2018).

55. Tahamtan, I. & Bornmann, L. What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. *Scientometrics* **121**, 1635–1684 (2019).

56. Bhattacharya, J. & Packalen, M. *Stagnation and Scientific Incentives (Working Paper 26752)*, https://www.nber.org/papers/w26752 (2020).

57. Azoulay, P., Graff Zivin, J. S. & Manso, G. Incentives and creativity: evidence from the academic life sciences. *RAND J. Econ.* **42**, 527–554 (2011).

58. Baltimore, D. Viral RNA-dependent DNA polymerase: RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**, 1209–1211 (1970).

59. Page, L. Method for node ranking in a linked database. US patent 6,285,999 (2001).

60. Axel, R., Wigler, M. H. & Silverstein, S. J. Processes for inserting DNA into eucaryotic cells and for producing proteinaceous materials. US patent 4,634,665 (1983).

61. Hawbaker, M. S. Soybean variety SE90346. US patent 6,958,436 (2005).

62. Katsuki, T. & Sharpless, K. B. The first practical method for asymmetric epoxidation. *J. Am. Chem. Soc.* **102**, 5974–5976 (1980).

63. Riess, A. G., et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron. J.* **116**, 1009

(1998).

64. Dirac, P. A. M. The quantum theory of the electron. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **117**, 610–624 (1928).

65. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977).

66. Bednorz, J. G. & Müller, K. A. Possible high T_c superconductivity in the Ba-La-Cu-O system. *Z. Phys. B Condens. Matter* **64**, 189–193 (1986).

67. Wuchty, S., Jones, B. F. & Uzzi, B. The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).

68. Guimera, R., Uzzi, B., Spiro, J. & Amaral, L. A. N. Team assembly mechanisms determine collaboration network structure and team performance. *Science* **308**, 697–702 (2005).

69. Jones, B. F., Wuchty, S. & Uzzi, B. Multi-university research teams: shifting impact, geography, and stratification in science. *Science* **322**, 1259–1262 (2008).

70. Grömping, U. Estimators of relative importance in linear regression based on variance decomposition. *Am. Stat.* **61**, 139–147 (2007).

71. Mukherjee, S., Uzzi, B., Jones, B. & Stringer, M. A new method for identifying recombinations of existing knowledge associated with high-impact innovation. *J. Prod. Innov. Manage.* **33**, 224–236 (2016).

72. Christianson, N. H., Sizemore Blevins, A. & Bassett, D. S. Architecture and evolution of semantic networks in mathematics texts. *Proc. R. Soc. A* **476**, 20190741 (2020).

73. Newman, M. E. J. The structure of scientific collaboration networks. *Proc. Natl Acad. Sci. USA* **98**, 404–409 (2001).

74. Newman, M. E. J. Scientific collaboration networks. I. Network construction and fundamental results. *Phys. Rev. E* **64**, 016131 (2001).

75. Uzzi, B. & Spiro, J. Collaboration and creativity: the small world problem. *Am. J. Sociol.* **111**, 447–504 (2005).

76. Funk, R. J. Making the most of where you are: geography, networks, and innovation in organizations. *Acad. Manage. J.* **57**, 193–222 (2014).

77. Barabási, A.-L. *Network Science* (Cambridge Univ. Press, 2016).

78. Blau, D. M. & Weinberg, B. A. Why the US science and engineering workforce is aging rapidly. *Proc. Natl Acad. Sci. USA* **114**, 3879–3884 (2017).

79. Cui, H., Wu, L. & Evans, J. A. Aging scientists and slowed advance. Preprint at https://doi.org/10.48550/arXiv.2202.04044 (2022).

80. Azoulay, P., Fons-Rosen, C. & Graff Zivin, J. S. Does science advance one funeral at a time? *Am. Econ. Rev.* **109**, 2889–2920 (2019).

Acknowledgements

This study was supported by the National Science Foundation (grant Nos. 1829168, 1932596 and 1829302).

Author information

Authors and Affiliations

1. Carlson School of Management, University of Minnesota, Minneapolis, MN, USA

   Michael Park & Russell J. Funk

2. School of Sociology, University of Arizona, Tucson, AZ, USA

   Erin Leahey

Contributions

R.J.F. and E.L. collaboratively contributed to the conception and design of the study. R.J.F. and M.P. collaboratively contributed to the acquisition, analysis and interpretation of the data. R.J.F. created software used in the study. R.J.F., E.L. and M.P. collaboratively drafted and revised the manuscript.

Corresponding author

Correspondence to Russell J. Funk.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Diana Hicks and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Distribution of CD_5.

This figure gives an overview of the distribution of CD_5 for papers (n = 24,659,076) and patents (n = 3,912,353). Panels **a** and **c** show counts of papers and patents over discrete intervals of CD_5. Panels **b** and **d** show the distribution of CD_5 over time, within 10 (papers) and 5 (patents) year intervals, using letter-value plots. These plots are similar to boxplots, but generally provide more reliable summaries for large datasets. They are drawn by identifying the median of the underlying distribution and then recursively drawing boxes outward from there in either direction that encompass half of the remaining data.

Source data

Extended Data Fig. 2 CD index measured using alternative forward citation windows.

This figure evaluates the sensitivity of our results to the use of different forward citation windows when computing the CD index for papers (n = 24,659,076) and patents (n = 3,912,353). In the main text, the index is computed based on citations made to papers and patents and their backward references as of 5 years after the year of publication. **a** and **c** plot the CD index using a longer, 10 year forward window, for papers and patents, respectively. **b** and **d** plot the CD index using all forward citations made to sample papers and patents as of the year 2017. Shaded bands correspond to 95% confidence intervals. Overall, the results mirror those reported in the main text, although the decline is somewhat steeper using longer forward citation windows, suggesting our primary results may represent a more conservative estimate.

Source data

Extended Data Fig. 3 Diversity of language use in science and technology over time.

This figure shows changes in the ratio of unique to total words (also known as the type-token ratio) over time based on data from the abstracts of papers

(**a**, n = 76 WoS research area × year observations) and patents (**b**, n = 229 NBER technology category × year observations). For papers, lines correspond to WoS research areas; for patents, lines correspond to NBER technology categories. For paper abstracts, lines begin in 1992 because WoS does not reliably record abstracts for papers published prior to the early 1990s. The ratio of unique to total words is computed separately by field (i.e., the uniqueness of words and total word counts are determined within WoS research areas and NBER technology categories). If disruption is decreasing, we may plausibly expect to see a decrease in the diversity of words used by scientists and inventors, as discoveries and inventions will be less likely to create departures from the status quo, and will therefore be less likely to need to introduce new terminology. For both papers and patents, we observe declining diversity in word use over time, which is consistent with this expectation and corroborates our findings using the CD index.

Source data

Extended Data Fig. 4 Declining combinatorial novelty.

This figure shows changing patterns in the combinatorial novelty/conventionality of papers (**a**, n = 24,659,076) and patents (**b**, n = 3,912,353), using a previously proposed measure of “atypical combinations”[14]. The measure quantifies the degree to which the prior work cited by a paper or patent would be expected by chance. For papers, we follow prior work[14] and consider combinations of cited journals. If a paper made three citations to prior work, and that work was published in three different journals—*Nature*, *Cell*, and *Science*—then there are three combinations—*Nature* × *Cell*, *Nature* × *Science*, and *Science* × *Cell*. To determine the degree to which each combination would be expected by chance, the frequency of observed pairings is compared to those in 10 “rewired” copies of the overall citation network, using a z-score. For patents, there is no natural analogue to journals, and therefore we consider pairings of primary United States Patent Classification (USPC) system codes. We present the results of this analysis following the approach of prior work[14], which plots the cumulative distribution function of the measure. In general, there is a rightward shift in the cumulative distributions over time, suggesting

that for both papers and patents, combinations are more conventional than would be expected by chance, consistent with what we would anticipate based on our results using the CD index. For patents, there is also a smaller shift in the opposite direction on the left side of the distribution, suggesting that novel patents in recent decades are somewhat more novel than novel patents in earlier decades. Overall, however, the bulk of the distribution is moving rightward, indicating greater conventionality.

Extended Data Fig. 5 Contribution of field, year, and author effects.

This figure shows the relative contribution of field, year, and author fixed effects to the adjusted R^2 in regression models predicting CD_5. The top bar shows the results for papers (n = 80,607,091 paper × author observations); the bottom bar shows the results for patents (n = 8,319,826 patent × inventor observations). The results suggest that for both papers and patents, stable characteristics of authors contribute significantly to patterns of disruptiveness. Moreover, relatively little of the variation is accounted for by field-specific factors.

Source data

Extended Data Fig. 6 CD index over time across data sources.

This figure shows changes in CD_5 over time across four additional data sources (the WoS [n = 24,659,076] and Patents View [n = 3,912,353] lines are included for reference): JSTOR (n = 1,703,353), the American Physical Society corpus (n = 478,373), Microsoft Academic Graph (n = 1,000,000), and PubMed (n = 16,774,282). Colours indicate the six different data sources. Shaded bands correspond to 95% confidence intervals. The figure indicates that the decline in disruption is unlikely to be driven by our sample choice of WoS papers and Patents View patents.

Source data

Extended Data Fig. 7 Alternative measures of disruption.

This figure shows the decline in the disruption of papers (**a**, n = 100,000) and patents (**b**, n = 100,000) based on two alternative measures of disruption. The blue lines calculate disruption using a measure proposed in Bornmann et al.[13], DI_{l}^{nok} where l = 5, which makes the measure more resilient to marginal changes in the number of papers or patents that only cite the focal work's references. The orange lines calculate disruption using a measure proposed in Leydesdorff et al.[15], DI*, which makes the measure less sensitive to small changes in the forward citation patterns of papers or patents that make no backward citations. Shaded bands correspond to 95% confidence intervals. With both alternative measures, we observe decreases in disruption for papers and patents, suggesting that the decline is not an artefact of our operationalization of disruption.

Source data

Extended Data Fig. 8 Robustness to changes in publication, citation, and authorship practices.

This figure evaluates whether declines in disruptiveness may be attributable to changes in publication, citation, and authorship practices for papers (n = 24,659,076) and patents (n = 3,912,353). Panels **a** and **d** adjust for these changes using a normalization approach. We present two alternative versions of the CD index, both of which account for the tendency for papers and patents to cite more prior work over time. Blue lines indicate normalization at the paper level (accounting for the number of citations made by the focal paper/patent). Orange lines indicate normalization at the field and year level (accounting for the mean number of citations made by papers/patents in the focal field and year). Panels **b** (papers) and **e** (patents) adjust for changes in publication, citation, and authorship practices using a regression approach. The panels show predicted values of CD_5 based on regressions reported in Models 4 (papers) and 8 (patents) of Supplementary Table 1, which adjust for field × year—*Number of new papers/patents*, *Mean number of papers/patents*

cited, *Mean number of authors/inventors per paper/patent*—and paper/patent-level—*Number of papers/patents cited*, *Number of unlinked references*—characteristics. Predictions are made separately for each year indicator included in the models; we then connect these separate predictions with lines to aid interpretation. Finally, Panels **c** (papers) and **f** (patents) adjust for changes in publication, citation, and authorship practices using a simulation approach. The panels plot z-scores that compare values of CD_5 obtained from the observed citation networks to those obtained from randomly rewired copies of the observed networks. Across all six panels, shaded bands correspond to 95% confidence intervals.

Source data

Extended Data Fig. 9 Growth of scientific and technological knowledge.

This figure shows the number of papers (n = 24,659,076) published (**a**) and patents (n = 3,912,353) granted (**b**) over time. For papers, lines correspond to WoS research areas; for patents, lines correspond to NBER technology categories.

Source data

Extended Data Table 1 Regression models of disruptiveness and the use of prior knowledge

Supplementary information

Supplementary Information

Supplementary Sections 1–3, Tables 1–3 and References.

Reporting summary

Peer Review File

Source data

[Source Data Fig. 2.]

[Source Data Fig. 3.]

[Source Data Fig. 4.]

[Source Data Fig. 5.]

[Source Data Fig. 6.]

[Source Data Extended Data Fig. 1.]

[Source Data Extended Data Fig. 2.]

[Source Data Extended Data Fig. 3.]

[Source Data Extended Data Fig. 5.]

[Source Data Extended Data Fig. 6.]

[Source Data Extended Data Fig. 7.]

[Source Data Extended Data Fig. 8.]

[Source Data Extended Data Fig. 9.]

Rights and permissions

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Reprints and Permissions

About this article

Cite this article

Park, M., Leahey, E. & Funk, R.J. Papers and patents are becoming less disruptive over time. *Nature* **613**, 138–144 (2023). https://doi.org/10.1038/s41586-022-05543-x

- Received: 14 February 2022
- Accepted: 08 November 2022
- Published: 04 January 2023
- Issue Date: 05 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05543-x

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

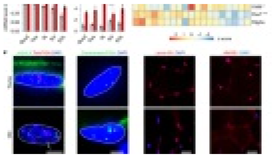
‘Disruptive’ science has declined — and no one knows why

- Max Kozlov

Nature News 04 Jan 2023

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05543-x

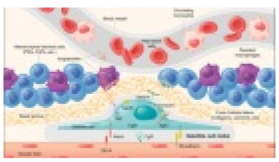
| Section menu | Main menu |



Characterization of cellular senescence in aging skeletal muscle

15 July 2022

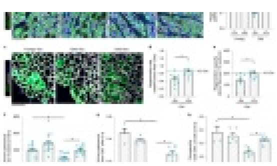
Xu Zhang, Leena Habiballa, … Nathan K. LeBrasseur



Skeletal muscle stem cells in comfort and stress

21 December 2018

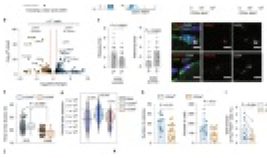
Brendan Evano & Shahragim Tajbakhsh



Exercise rejuvenates quiescent skeletal muscle stem cells in old mice through restoration of Cyclin D1

13 April 2020

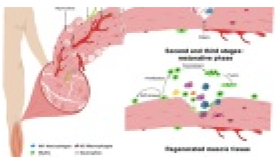
Jamie O. Brett, Marina Arjona, … Thomas A. Rando



[FoxO maintains a genuine muscle stem-cell quiescent state until geriatric age]

26 October 2020

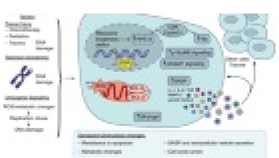
Laura García-Prat, Eusebio Perdiguero, … Pura Muñoz-Cánoves



[The jam session between muscle stem cells and the extracellular matrix in the tissue microenvironment]

17 February 2022

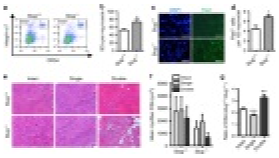
Mafalda Loreti & Alessandra Sacco



[Cellular senescence in musculoskeletal homeostasis, diseases, and regeneration]

10 September 2021

Mei Wan, Elise F. Gray-Gaillard & Jennifer H. Elisseeff



[The transcription factor Slug represses p16Ink4a and regulates murine muscle stem cell aging]

12 June 2019

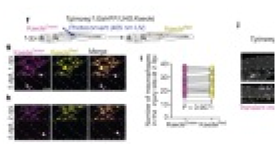
Pei Zhu, Chunping Zhang, … Wen-Shu Wu



[Control of satellite cell function in muscle regeneration and its disruption in ageing]

18 October 2021

Pedro Sousa-Victor, Laura García-Prat & Pura Muñoz-Cánoves



[Macrophages provide a transient muscle stem cell niche via NAMPT secretion]

10 February 2021

Dhanushika Ratnayake, Phong D. Nguyen, … Peter D. Currie

- Article

- Open Access
- Published: 21 December 2022

Senescence atlas reveals an aged-like inflamed niche that blunts muscle regeneration

- Victoria Moiseeva,
- Andrés Cisneros,
- Valentina Sica,
- Oleg Deryagin,
- Yiwei Lai,
- Sascha Jung,
- Eva Andrés,
- Juan An,
- Jessica Segalés,
- Laura Ortet,
- Vera Lukesova,
- Giacomo Volpe,
- Alberto Benguria,
- Ana Dopazo,
- Salvador Aznar-Benitah,
- Yasuteru Urano,
- Antonio del Sol,
- Miguel A. Esteban,
- Yasuyuki Ohkawa,
- Antonio L. Serrano,
- Eusebio Perdiguero &
- …
- Pura Muñoz-Cánoves

Show authors

Nature **volume 613**, pages 169–178 (2023)

- 24k Accesses
- 281 Altmetric
- Metrics details

Abstract

Tissue regeneration requires coordination between resident stem cells and local niche cells[1,2]. Here we identify that senescent cells are integral components of the skeletal muscle regenerative niche that repress regeneration at all stages of life. The technical limitation of senescent-cell scarcity[3] was overcome by combining single-cell transcriptomics and a senescent-cell enrichment sorting protocol. We identified and isolated different senescent cell types from damaged muscles of young and old mice. Deeper transcriptome, chromatin and pathway analyses revealed conservation of cell identity traits as well as two universal senescence hallmarks (inflammation and fibrosis) across cell type, regeneration time and ageing. Senescent cells create an aged-like inflamed niche that mirrors inflammation associated with ageing (inflammageing[4]) and arrests stem cell proliferation and regeneration. Reducing the burden of senescent cells, or reducing their inflammatory secretome through CD36 neutralization, accelerates regeneration in young and old mice. By contrast, transplantation of senescent cells delays regeneration. Our results provide a technique for isolating in vivo senescent cells, define a senescence blueprint for muscle, and uncover unproductive functional interactions between senescent cells and stem cells in regenerative niches that can be overcome. As senescent cells also accumulate in human muscles, our findings open potential paths for improving muscle repair throughout life.

Main

Tissue regeneration is necessary for life. Its success requires coordinated communication between resident stem cells and the surrounding niche cells. Examples can be found in the crosstalk between skeletal muscle stem cells

(called satellite cells (SCs)) and macrophages or fibroadipogenic progenitors (FAPs; also called mesenchymal cells) (reviewed in refs. [1,2]). However, the identities of the niche cellular components and secreted factors that regulate tissue regeneration are not fully characterized. During ageing, tissue regenerative functions decline, in part due to stem cell-intrinsic accumulation of damage (for example, DNA damage, loss of proteostasis and oxidative damage), functional decay of niche cells and increased inflammation (inflammageing)[4,5,6,7,8,9] through uncharacterized mechanisms.

Cellular senescence is a state of durable cell-cycle arrest of dysfunctional cells, which acquire a bioactive senescence-associated secretory phenotype (SASP)[10]. Senescent cells accumulate with ageing and age-associated diseases, limiting lifespan and healthspan in mice[11,12,13,14,15,16]. These cells also accumulate after irradiation and impair muscle function[17]. By contrast, senescent cells have shown beneficial effects as tumour suppressors, during embryo development, and in liver and skin repair[18,19,20,21] or reprogramming[22,23,24]. However, little is known about the identity of senescent cells in vivo or their contribution to tissue regeneration owing to their scarcity and presumed heterogeneity, and the lack of universal senescence markers and isolation methods[3].

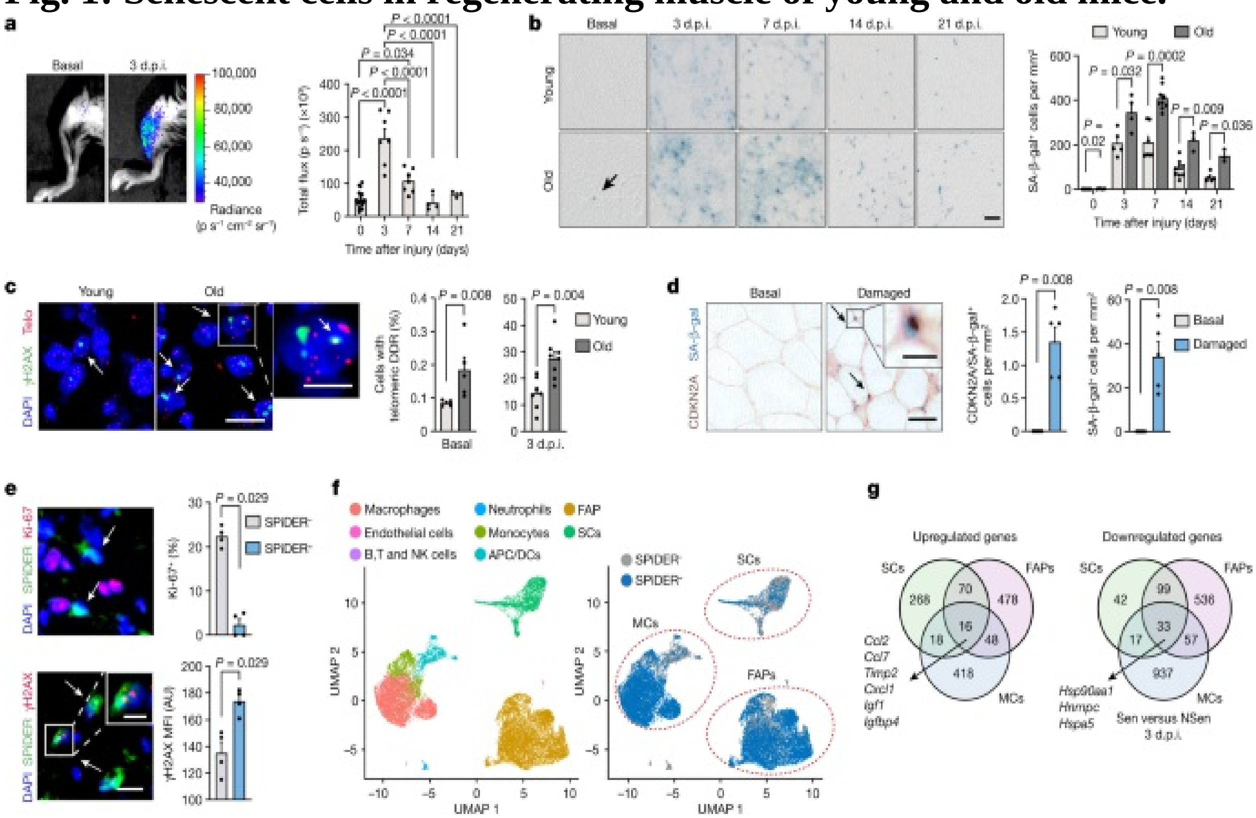
Here we isolated and defined senescent cells in vivo at the molecular and functional levels. We demonstrate that senescent cells are integral regenerative niche components that repress regeneration at all stages of life. Using single-cell transcriptomics, a method to separate senescent from non-senescent subsets of various niche cell types, and deep transcriptomics and chromatin analyses, we produced a lifetime atlas of in vivo senescent cells in regenerating skeletal muscle.

The injured niche gains senescent cells

To study cellular senescence in skeletal muscle, we used p16-3MR mice, which express *Renilla* luciferase, monomeric red fluorescent protein (RFP) and a viral thymidine kinase under the *Cdkn2a* (also known as $p16^{INK4a}$) promoter[18], and analysed luciferase activity (a proxy of CDKN2A-expressing senescent cells). We also used wild-type (WT) mice and analysed

senescence-associated beta-galactosidase (SA-β-gal) activity as a marker of senescent cells. We induced injury by intramuscular injection of cardiotoxin (CTX) into either young mice (aged 3–6 months) or very old mice (28-month-old mice and older (geriatric age)). Neither luciferase nor SA-β-gal activities were detectable in resting muscles, but were substantially induced at 3 days post-injury (d.p.i.) and were still elevated at 7 d.p.i., after which they decreased (Fig. 1a,b and Extended Data Fig. 1a). Senescent cells were more abundant and persisted longer in injured muscles of old mice (Fig. 1b and Extended Data Fig. 1a), correlating with their less efficient recovery process[5,6] (Extended Data Fig. 1b). Moreover, the telomeric DNA-damage response (DDR) (γH2AX telomere immunostaining) was also higher in regenerating muscle from old compared with from young mice or in non-injured (basal) muscle (Fig. 1c). Additional markers confirmed that injury induced a transient accumulation of senescent cells in both age groups (Extended Data Fig. 1c). Similar to in mouse muscle, SA-β-gal- and CDKN2A-positive cells were present in biopsy samples of damaged adult human muscles (Fig. 1d). Thus, senescent cells are induced in the damaged muscle of both humans and mice.

Fig. 1: Senescent cells in regenerating muscle of young and old mice.



a–**e**, Representative images and quantification. **a**, Luciferase activity in the muscles of young (3–6 months) p16-3MR mice at the indicated d.p.i. with CTX. p s^{-1}, photons per second. *n* = 18 muscles from 12 mice (basal), 7 muscles from 5 mice (3 d.p.i.), 7 muscles from 6 mice (7 d.p.i.) and 4 muscles from 3 mice (14 d.p.i. and 21 d.p.i.). **b**, SA-β-gal$^+$ cells in the TA of young mice (*n* = 5 (0, 3 and 21 d.p.i.), *n* = 7 (7 d.p.i.) and *n* = 9 (14 d.p.i.) mice) or old mice (aged more than 28 months; *n* = 4 (0 and 3 d.p.i.), *n* = 9 (7 d.p.i.) and *n* = 3 (14 and 21 d.p.i.) mice). **c**, cells with telomeric DDR (telo) in basal and 3 d.p.i. TA from young mice (*n* = 5 (basal) and *n* = 7 (3 d.p.i.) mice) and old mice (*n* = 5 (basal) and *n* = 8 (3 d.p.i.) mice). **d**, SA-β-gal and CDKN2A$^+$ immunohistochemistry of uninjured or damaged human muscle samples (*n* = 5 samples per group; aged 81 ± 7.5 years). The arrows show double-positive cells. **e**, Ki-67 positivity (*n* = 4 mice per group) and γH2AX intensity (*n* = 4 mice per group) in SPiDER$^+$ and SPiDER$^-$ cells in 4 d.p.i. TA. AU, arbitrary units; MFI, mean fluorescence intensity. **f**, Complete 21,449-cell transcriptomic atlas from SPiDER$^+$ and SPiDER$^-$ samples of 3 d.p.i. young muscles (right). Data are shown as a uniform manifold approximation and projection (UMAP) to visualize variation in single-cell transcriptomes. Unsupervised clustering resolved at least eight cell types (colour coded). Left, SPiDER$^+$ cells ascribed to main populations. APC, antigen-presenting cell; DCs, dendritic cells; NK, natural killer. **g**, The overlap between differentially expressed genes in SCs, FAPs and MCs at 3 d.p.i. from scRNA-seq (Sen versus NSen, false-discovery rate (FDR) < 0.05). Scale bars, 50 μm (**b** and **d**, low magnification), 10 μm (**c** and **e**, low magnification) and 5 μm (**c**–**e**, high magnification). For **a**–**e**, data are mean ± s.e.m. *P* values were calculated using Tukey's tests (**a**) and Mann–Whitney *U*-tests (**b**–**e**).

Source data

To determine the nature of senescent cells in vivo, we established a fluorescence-activated cell sorting (FACS) method based on the fluorescent probe SPiDER-β-gal (hereafter, SPiDER; similar to the C12FDG fluorescent probe)[25,26], which labels all SA-β-gal$^+$ cells (Extended Data Fig. 1d). No cells were labelled by SPiDER in unperturbed muscle. By contrast, in injured muscles, cells with a high SPiDER signal (SPiDER$^+$) and SPiDER$^-$ cells were detected (Extended Data Fig. 1e). The SPiDER$^+$ fraction contained cells

that were highly enriched for all of the tested senescence markers (such as SA-β-gal activity, larger cell size, lamin B1 reduction, cell-cycle arrest and increased DNA damage) (Fig. 1e and Extended Data Fig. 1f). Scoring for four of these markers, the SPiDER^{+} cell fraction was 93.6% enriched in senescent cells, demonstrating the specificity of the separation method (Extended Data Fig. 1g).

We next aimed to generate a single-cell RNA-sequencing (scRNA-seq) atlas of senescent cells from regenerating muscle in an unbiased manner. To control for cell-type differences in autofluorescence, and to obtain potential minor, non-haematopoietic senescent cells, we sorted SPiDER^{+} and SPiDER^{-} fractions from CD45^{+} and CD45^{-} populations and merged them after scRNA-seq (Extended Data Fig. 1h). The SPiDER^{+} populations comprised predominantly monocytes and macrophages (hereafter, myeloid cells (MCs)), FAPs, and SCs and their progeny, while smaller populations included antigen-presenting cells, neutrophils, endothelial cells, and B, T, and natural killer cells (Fig. 1f). Differential expression analysis revealed a core signature of 16 upregulated and 33 downregulated genes in the three major senescent populations (Fig. 1g), of which the former included inflammatory and matrix-remodelling/fibrotic SASP factors (such as *Ccl2*, *Ccl7*, *Igfbp4* and *Timp2*).

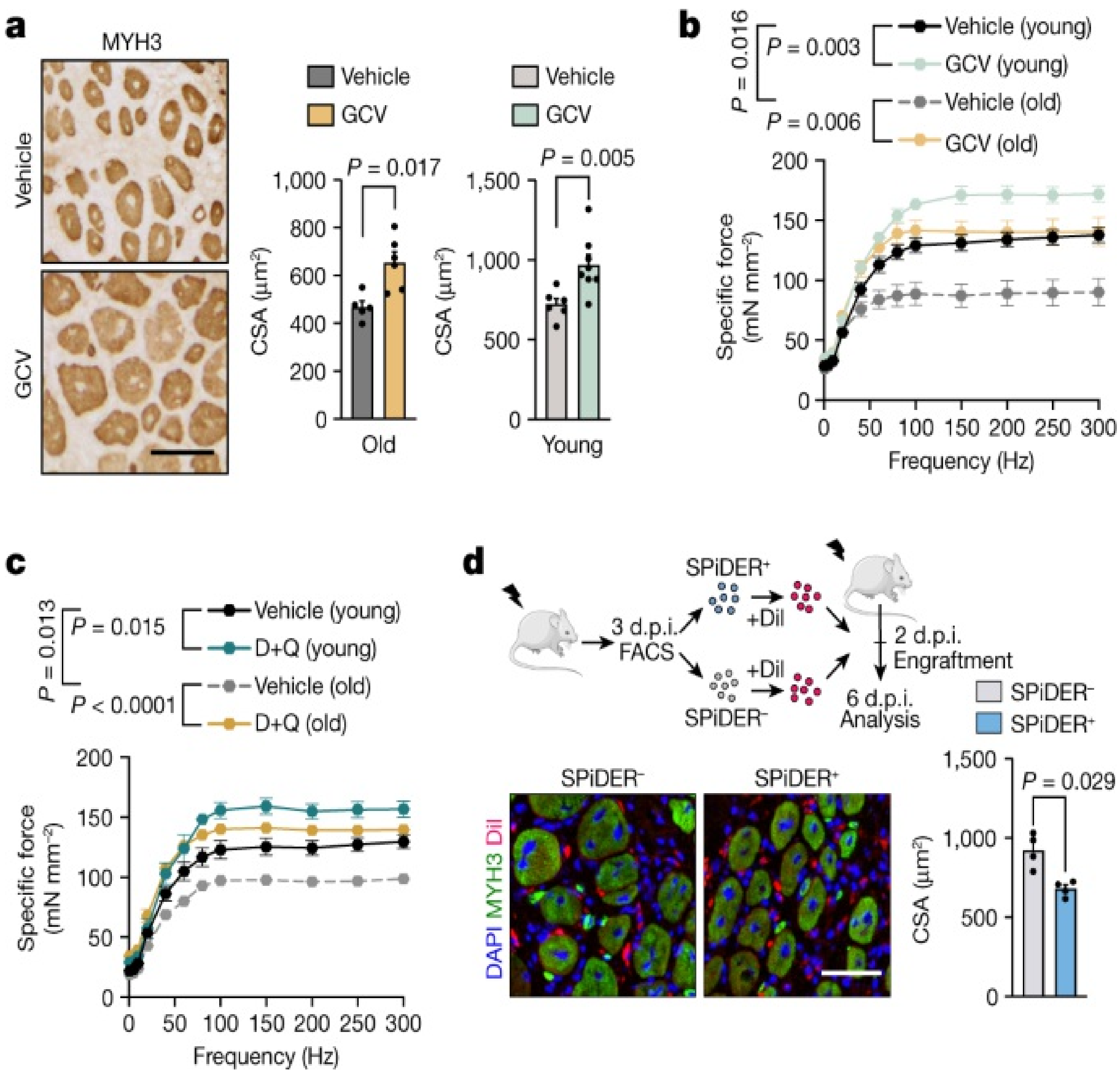
Immunostaining of senescence markers (CDKN2A and γH2AX) labelled cells close to the three major niche cell types identified by scRNA-seq (for example, PAX7^{+} for SCs; PDGFRα^{+} for FAPs and F4/80^{+} for MCs) (Extended Data Fig. 1i) in regenerating mouse tissue. Furthermore, markers for SCs, FAPs and MCs co-localized with senescence markers (in the order MCs ≥ FAPs > SCs) (Extended Data Fig. 1j). Importantly, damaged human muscle also contained SCs (PAX7^{+}), FAPs (PDGFRα^{+}) and MCs (CD68^{+}) positive for the DDR marker 53BP1 (Extended Data Fig. 1k). Together, these results provide: (1) a strategy to isolate senescent cells from tissue that does not rely on transgenic mice; (2) a single-cell cartography of senescent cells in vivo; and (3) the identity of new senescent-cell niche constituents after injury.

Senescence hampers muscle regeneration

To determine the role of senescent cells in muscle regeneration, we treated young and old p16-3MR mice daily with ganciclovir (GCV) during this

process. GCV reduced the presence of senescent cells in injured muscle, indexed by lower luciferase activity and number of SA-β-gal^{+}cells (Extended Data Fig. 2a, b). GCV treatment rescued defective muscle regeneration, reduced inflammation and fibrosis, and enhanced force generation in old mice, and accelerated the regenerative process in young p16-3MR mice (Fig. 2a,b and Extended Data Fig. 2c–f). Daily treatment with the senolytic compounds dasatinib and quercetin (hereafter D+Q)[15] also reduced SA-β-gal^{+} cell number, accelerated regeneration, increased force, and reduced matrix deposition and inflammation at both ages (Fig. 2c and Extended Data Fig. 2g–k). Similar benefits were obtained when GCV or D+Q were transiently administered (starting at 3 d.p.i.) (Extended Data Fig. 2l,m). Consistently, transplantation of sorted SPiDER^{+} (senescent) and SPiDER^{-} (non-senescent) cells, labelled with the lipophilic vital dye DiI, into preinjured muscle revealed that only SPiDER^{+} cells delayed regeneration of young host muscles (Fig. 2d). Thus, senescent cells were detrimental to muscle regeneration in both young and old mice.

Fig. 2: Cellular senescence hampers muscle regeneration throughout life.



a, Representative images and quantification of the cross-sectional area (CSA) of embryonic myosin heavy chain (MYH3)-positive fibres in regenerating TA from vehicle- or GCV-treated old (n = 5 TA from 3 mice (vehicle) and n = 6 TA from 3 mice (GCV)) and young (n = 6 TA from 4 mice (vehicle) and n = 8 TA from 4 mice (GCV)) p16-3MR mice at 7 d.p.i. **b**, Specific force–frequency curves of EDL muscles of vehicle- or GCV-treated young (n = 11 EDL from 7 mice (vehicle) and n = 9 EDL from 7 mice (GCV)) and old (n = 5 EDL from 4 mice (vehicle) and n = 7 EDL from 5 mice (GCV)) p16-3MR mice at 10 d.p.i. **c**, Specific force–frequency curves as described in **b**, but for vehicle- and D+Q-treated young (n = 8 EDL from 5 mice (vehicle) and n = 7

EDL from 5 mice (D+Q)) and old (*n* = 5 EDL from 3 mice (vehicle) and *n* = 7 EDL from 5 mice (D+Q group)) mice at 10 d.p.i. **d**, An equal number of SPiDER$^+$ or SPiDER$^-$ cells from young 3 d.p.i. regenerating muscles were stained with Dil and transplanted into the preinjured TA of young recipient mice for 4 days. *n* = 4 mice per group. Strategy schematic, representative images and quantification of the CSA of MYH3$^+$ fibres are shown. The schematic in **d** was created using Servier Medical Art, CC BY 3.0. Scale bars, 50 µm (**a** and **d**). For **a**–**d**, data are mean ± s.e.m. *P* values were calculated using two-way analysis of variance (ANOVA) and mixed-effects analysis (**b** and **c**) and Mann–Whitney *U*-tests (**a** and **d**).

Source data

Senescence in acute and chronic damage

To exclude that the observed roles of senescent cells depended on the degree of tissue injury, we (1) induced transient, mild injury with localized muscle micropunctures; and (2) examined the chronic, severely injured muscle of mdx mice (a model of Duchenne muscular dystrophy)[27]. Muscle micropunctures induced acute and transient SA-β-gal$^+$ cell presence and luminescence activity in WT and p16-3MR mice, respectively (Extended Data Fig. 3a,b). Reducing the accumulation of senescent cells in p16-3MR mice (by GCV) or WT mice (by D+Q) increased the size of regenerating myofibres and decreased inflammation and matrix deposition (Extended Data Fig. 3c–e). Next, p16-3MR mice were brought to the mdx background. Senescent cells were more numerous and persistent in the muscles of mdx/p16-3MR mice compared with in the non-dystrophic muscles of p16-3MR mice (Extended Data Fig. 3f–h), in agreement with chronic damage in mdx muscles (Extended Data Fig. 3i). Longer-term treatment (twice weekly for 2 months) of mdx/p16-3MR mice with GCV or mdx mice with D+Q reduced the burden of senescent cells and alleviated the regenerative impairment, as shown by larger myofibres, lower fibrosis and enhanced muscle force (Extended Data Fig. 3j–m). Thus, irruption of senescent cells in the muscle niche, either transiently (in mild injury) or persistently (in chronic injury), was deleterious for regeneration, irrespective of age. These findings challenge the prevalent view that senescent cells are beneficial when transiently present after acute injury, particularly in young tissue (reviewed in

refs. [28,29]).

Heterogeneity of senescent cells in vivo

For a deeper characterization of the three main populations of senescent cells from regenerating muscle, we established a FACS method based on the coupling of cell-surface markers of SCs, FAPs and MCs to the SPiDER reagent (Extended Data Fig. 4a). The method provided simultaneous isolation of the SPiDER$^+$ and SPiDER$^-$ subpopulations (hereafter referred to as senescent (Sen) and non-senescent (NSen) cells) of the three cell types, as validated by immunostaining and gene expression (Extended Data Fig. 4b,c), and at single-cell level (Extended Data Fig. 4d). Each Sen subpopulation scored positively for six accepted senescence markers—namely, increased SA-β-gal activity, large cell size, decreased proliferation, reduced lamin B1, lack of apoptotic programmed death and high *Cdkn2a* expression (Extended Data Fig. 4e–j). Moreover, both 8-oxoguanine (an indicator of reactive oxygen species (ROS)-induced DNA damage) (Extended Data Fig. 4k) and telomeric DDR were increased in the Sen subpopulations (Extended Data Fig. 4l), confirming the high level of senescent-cell enrichment of the SPiDER-sorting protocol.

For in-depth transcriptomic analysis of the distinct cell populations, we performed bulk RNA-seq analysis of Sen and NSen SCs, FAPs and MCs at early (3 d.p.i.) and late (7 d.p.i.) phases of regeneration, and NSen cells from basal (contralateral) muscles from young or old mice (a total of 36 distinct in vivo conditions) (Extended Data Fig. 5a). Principal component analysis (PCA) of all of the datasets revealed that the transcriptomes primarily clustered by cell type, rather than cell state, regeneration timepoint or mouse age (Extended Data Fig. 5b). Within each cell type, there was clear segregation according to cell state, with a strong influence of senescence (Extended Data Fig. 5c).

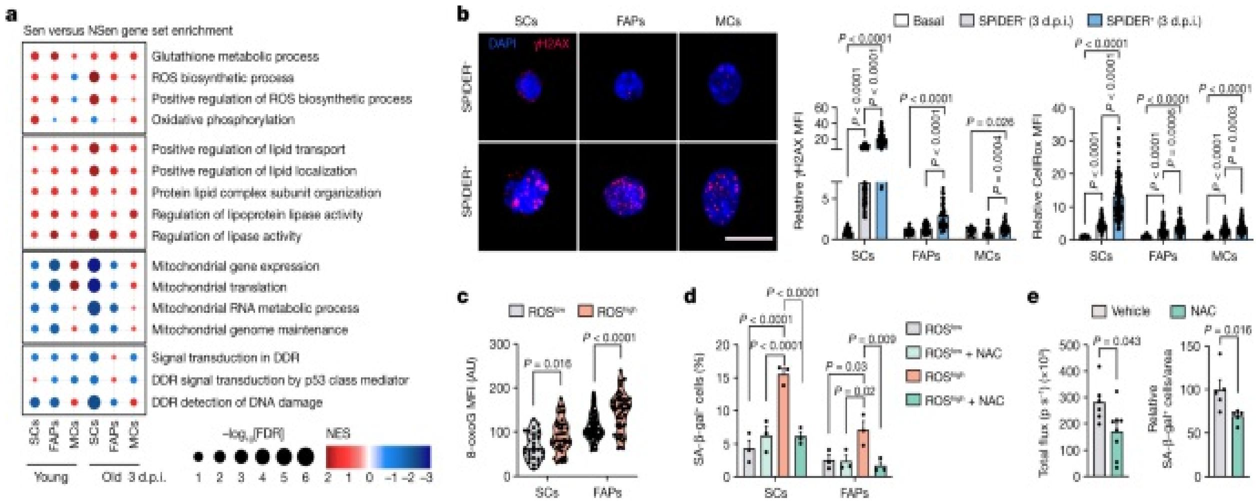
A comparison between the Sen and NSen subpopulations at 3 d.p.i. in young mice showed 2,251 differentially expressed genes in Sen SCs, 1,805 in Sen FAPs and 4,958 in Sen MCs with little overlap (Extended Data Fig. 5d,e). Many differentially expressed genes and enriched pathways were specific to each Sen cell type and revealed the cell of origin: muscle contraction and

integrin/cell-surface interactions in Sen SCs; actin cytoskeleton and elastic-fibre regulation in Sen FAPs; and immune functions and high lysosomal content in MCs (Extended Data Fig. 5f). Basic cell processes were downregulated in each Sen subpopulation (Extended Data Fig. 5f). However, with ageing, Sen cells gained alternative cell-fate traits, including macrophage-specific pathways and complement/coagulation cascades in Sen FAPs (Extended Data Fig. 6a); extracellular matrix organization and collagen formation in Sen MCs (Extended Data Fig. 6b); and immune features in Sen SCs (Extended Data Fig. 6c). All Sen cells from old mice gained further cell plasticity and pro-inflammatory traits (Extended Data Fig. 6d), which, in addition to their higher number (Fig. 1b,c and Extended Data Fig. 1a,c), may contribute to muscle regenerative failure at very old age (Extended Data Fig. 1b).

Tissue injury and ageing drive senescence

To understand how senescence is induced in the regenerative niche, we searched for pathways that are commonly enriched in all Sen cells early after injury (3 d.p.i.). Compared with basal cells, DNA-damage, cell-cycle regulation and inflammation pathways were upregulated, whereas homeostatic gene expression and protein translation pathways were downregulated in Sen cells (Extended Data Fig. 7a). Compared with NSen cells, Sen cells were enriched in pathways that are implicated in cellular stress, such as oxidative and metabolic stress (ROS and oxidative phosphorylation, and lipid transport and metabolism), with concomitant downregulation of DNA-damage-repair and mitochondrial-response pathways (Fig. 3a). All of the Sen cell types had more DNA-damage foci compared with NSen or basal cells (γH2AX immunostaining) (Fig. 3b), and ROS-induced DNA damage and telomeric DDR (Extended Data Fig. 4k,l). 3 d.p.i. Sen cells had more intense CellRox (ROS) staining compared with NSen (or basal) counterparts (Fig. 3b), which may relate to their mitochondrial dysfunction (Fig. 3a).

Fig. 3: Tissue injury and ageing prime niche cells for senescence through oxidative stress and DNA damage.



a, Common upregulated and downregulated gene sets (gene set enrichment analysis (GSEA), FDR < 0.25) related to the indicated functions in Sen versus NSen SCs, FAPs and MCs from young and old mice at 3 d.p.i. **b**, Representative images of γH2AX and quantification of γH2AX ($n = 24$ SCs (basal), 91 SCs (NSen), 78 SCs (Sen), 26 FAPs (basal), 34 FAPs (NSen), 35 FAPs (Sen), 20 MCs (basal), 22 MCs (NSen) and 29 MCs (Sen)) and CellRox levels ($n = 24$ SCs (basal), 95 SCs (NSen), 98 SCs (Sen), 26 FAPs (basal), 81 FAPs (NSen), 104 FAPs (Sen), 20 MCs (basal), 93 MCs (NSen) and 97 MCs (Sen)) in sorted SC, FAP and MC populations from basal and regenerating muscles at 3 d.p.i. **c**, Quantification of 8-oxoguanine (8-oxoG) levels in sorted ROS^{high} and ROS^{low} SCs and FAPs from young muscle at 1 d.p.i. ($n = 27$ SCs (ROS^{low}), $n = 40$ SCs (ROS^{high}), $n = 66$ FAPs (ROS^{low}) and $n = 50$ FAPs (ROS^{high})). **d**, ROS^{high} and ROS^{low} SCs and FAPs were isolated from regenerating muscle at 1 d.p.i., and were cultured in vitro for 3 days with or without NAC. Quantification of SA-β-gal^{+} cells in each population compared with basal cells. $n = 3$ mice per group. **e**, Young p16-3MR mice were injured with CTX and treated with NAC during regeneration. Left, *Renilla* luciferase activity in the TA at 4 d.p.i. $n = 6$ TA muscles from 6 mice (vehicle) and $n = 8$ TA muscles from 4 mice (NAC group). Right, quantification of SA-β-gal^{+} cells. $n = 5$ TA muscles from 4 mice in each group. For **b**, scale bar, 10 μm. For **b**, **d** and **e**, data are mean ± s.e.m. *P* values were calculated using Tukey's tests (**d**) and Mann–Whitney *U*-tests (**b**, **c** and **e**).

Source data

To assess the causality of the high ROS levels in driving injury-induced senescence, we sorted ROS^{high} and ROS^{low} SCs and FAPs from muscles at 1 d.p.i. (before the appearance of senescent cells). ROS^{high} cells sorted at 1 d.p.i. presented higher levels of 8-oxoguanine compared with ROS^{low} cells (Fig. 3c). Moreover, ROS^{high} cells, but not ROS^{low} cells, became senescent in culture, and ROS inhibition with the antioxidant *N*-acetylcysteine (NAC) blocked senescence entry of ROS^{high} cells (Fig. 3d). During regeneration, senescence was reduced by inhibiting ROS (Fig. 3e). Sen cells showed altered lipid-transport and lipoprotein-remodelling traits (Fig. 3a), consistent with the ability of oxidative stress to generate lipotoxicity that triggers inflammation and fibrosis[30].

To analyse the sole effect of ageing on resting-muscle cells, we compared basal cells of old and young mice and found upregulation of the immune-inflammatory response, DNA-damage and cell-cycle arrest, lipid metabolism, matrix remodelling and insulin-signalling genes, and downregulation of mitochondrial genes in the old mice (Extended Data Fig. 7b). Basal cells from old mice had more DNA-damage foci compared with their young counterparts (Fig. 1c). This coincided with higher ROS levels in cells of the old mice[6,9]. Matrix deposition (Extended Data Fig. 7c) and expression of cell-cycle inhibitors and inflammatory factors were also higher in resting muscles of old mice (Extended Data Fig. 7d), consistent with the concept of inflammageing[4]. Thus, injury and ageing led to the accumulation of stressors and activation of inflammatory/fibrotic pathways that primed old niche cells for senescence, resulting in a deeper senescent state after injury.

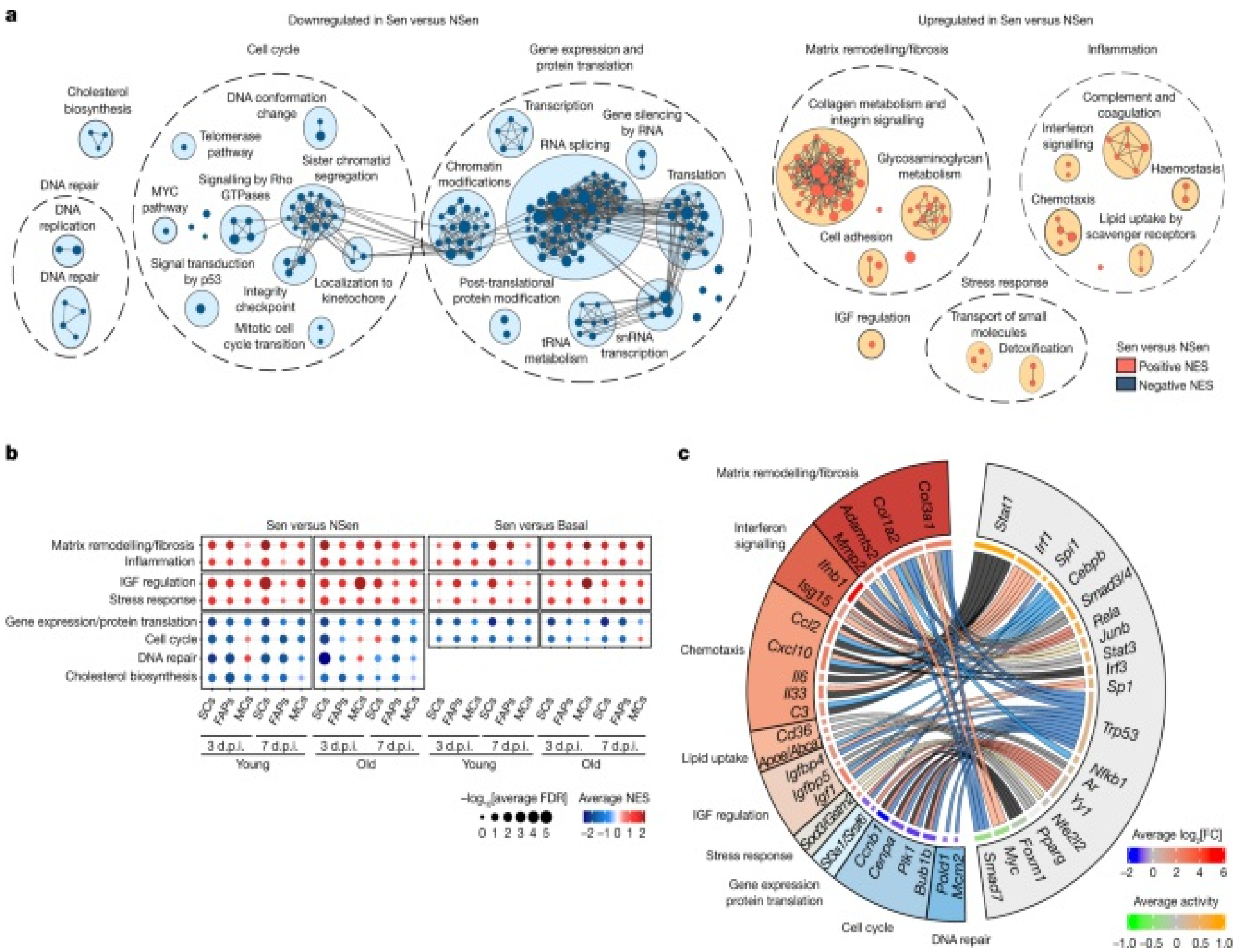
Two hallmarks define senescence in vivo

We next searched for commonly regulated traits in Sen cells across all ages and time points. Despite gene-expression heterogeneity (Extended Data Figs. 5e and 8a), 47 differentially expressed genes were largely conserved among conditions (Extended Data Fig. 8b), including pro-inflammatory cytokines (such as *Ccl2*, *Ccl7* and *Ccl8*), matrix/remodelling components (such as *Col1a2*, *Col3a1* and *Timp2*) and insulin growth factor (IGF) regulators (such as *Igfbp4*, *Igfbp6* and *Igfbp7*), that were previously linked to senescence. Expression of the pro-inflammatory/pro-fibrotic genes *Ccl2*, *Ccl7*, *Igf1*,

Igfbp4 and *Timp2* was also detected by scRNA-seq mapping (Fig. 1g). Key genes were validated using quantitative PCR with reverse transcription (RT–qPCR; Extended Data Fig. 8c).

Pathway-enrichment analysis showed upregulation of two major functions: inflammation (complement and coagulation, chemotaxis, interferon signalling and lipid uptake by scavenger receptors) and matrix remodelling/fibrosis (extracellular matrix glycoproteins and cell adhesion) (Fig. 4a). These two hallmarks are conserved in replicative senescence in vitro[31]. Minor conserved traits were related to stress responses (cell detoxification and small-molecule transport) and IGF regulation (IGF transport/uptake by IGFBPs) (Fig. 4a). By contrast, basic cell machinery processes were downregulated across conditions, with reduced gene expression, splicing, translation, mitotic cell-cycle processes, and DNA replication and repair functions (Fig. 4a). A comparison of Sen cells with their NSen and basal counterparts simultaneously showed that core senescence hallmarks were maintained (Fig. 4b and Extended Data Fig. 8d), excluding that they were due to the cell-growth arrest (quiescence) state per se.

Fig. 4: Two major common hallmarks define senescent cells across cell types, regeneration stage and lifespan.



a, Clusters of gene sets (GSEA) differentially enriched from Sen versus NSen SCs, FAPs and MCs from young or old mice at 3 or 7 d.p.i. Gene sets were considered to be common with FDR < 0.25 in at least 8 out of 12 comparisons. Node size is proportional to the number of genes identified in each gene set. The grey edges indicate gene overlap. **b**, Common clusters of gene sets (GSEA) from Sen versus NSen and Sen versus basal SCs, FAPs and MCs from young or old mice at 3 or 7 d.p.i. Gene sets were considered to be common with FDR < 0.25 in at least 8 out of 12 comparisons for Sen versus NSen and Sen versus basal. **c**, Chord diagram showing transcription factors that regulate the differentially expressed genes in Sen versus NSen and their categories. The green-to-orange scale indicates the average predicted transcription factor activity. The blue-to-red scale indicates the average base 2 logarithm fold change ($\log_2$[FC]) of a transcription factor target in Sen versus NSen cells. Chord width is proportional to the significance ($-\log_{10}$[FDR]) of canonical pathway and Gene Ontology

biological process (GO:BP) enrichment (gprofiler2) within a given functional category.

Source data

Chromatin analysis of senescent cells

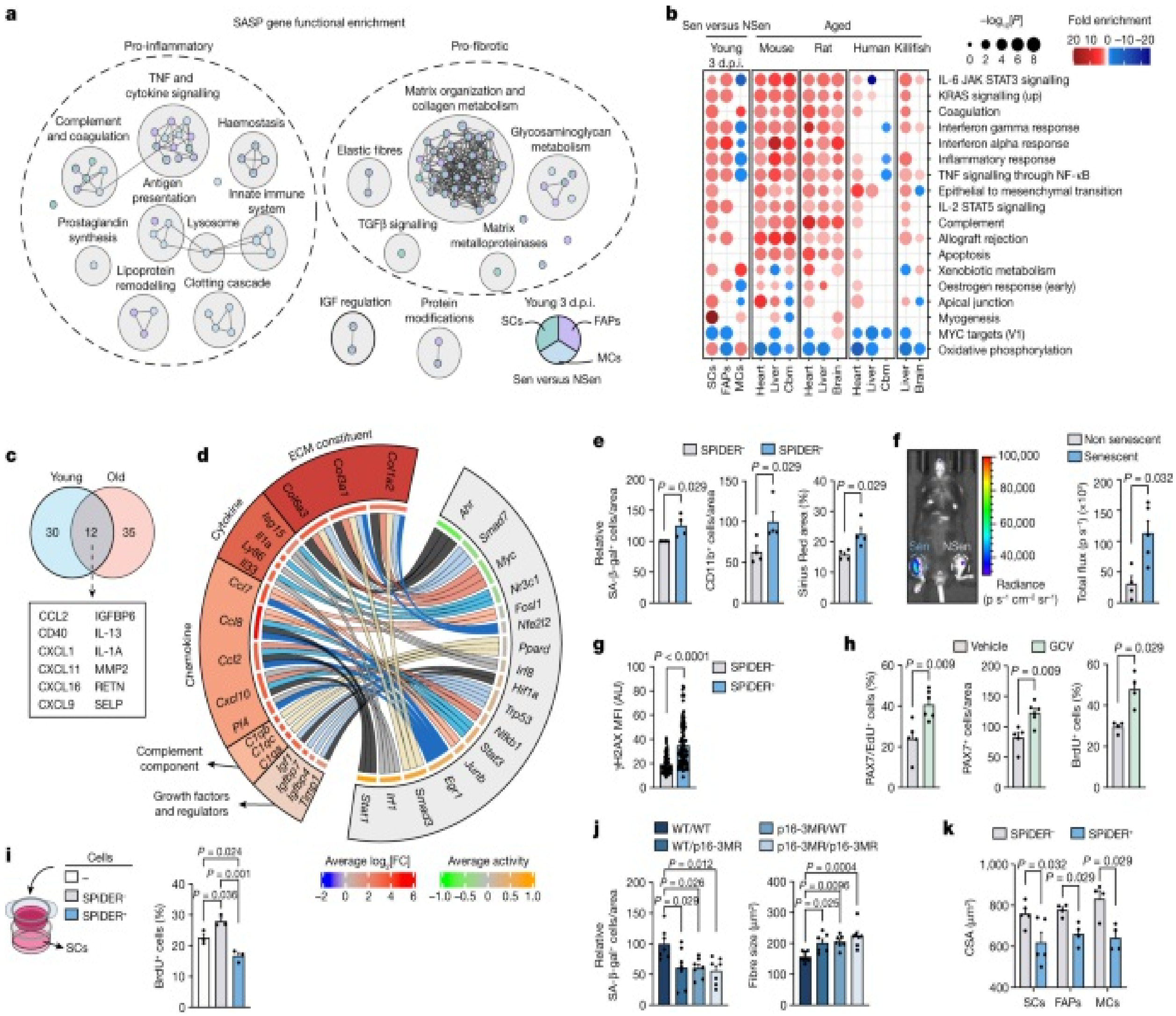
Transcription factor enrichment analysis in Sen cells in all conditions revealed enrichment occupancy of key regulators of inflammation and the SASP, including NF-κB[32], C/EBPβ[33] and STAT1/3 (ref. [34]), and of matrix remodelling/fibrosis, including SMAD3/4 (and inhibition of SMAD7)[35] (Fig. 4c). We also performed transposase-accessible chromatin high-throughput sequencing (ATAC-seq) analysis of Sen and NSen FAPs, MCs and SCs at early and late regeneration phases (and in basal cells) from young and old mice, and analysed the promoter accessibility of the genes that define senescence in our transcriptomic data. By comparing the signal for each ATAC-seq peak, we observed total reduced chromatin accessibility in Sen cells (Extended Data Fig. 9a). In Sen cells of old mice, promoters with medium peaks were more accessible. As the low number of cells used for ATAC-seq limited its resolution, we restricted our analysis to evaluate the overall chromatin accessibility between Sen and NSen cells. We detected an increased peak signal at proximal promoters of the *Ifnb1* and *Ccl8* genes in Sen cells, correlating with increased expression (Extended Data Fig. 9b). Motif-enrichment analysis of ATAC-seq data validated the transcription factors predicted by RNA-seq (NF-κB, C/EBPβ, STAT1/3, SMAD3/4) (Extended Data Fig. 9c), reinforcing that senescent cells within the muscle niche were of an inflammatory/fibrotic nature, regardless of age or time after injury.

A senescent aged-like inflamed niche

To understand how senescent cells impair muscle regeneration, we studied their SASP after injury and during ageing. We selected differentially expressed genes with extracellular or secreted protein products in Sen versus NSen cells. Depending on cell type and conditions, the number of SASP components ranged from 78 to 363, highlighting SASP diversity (Extended Data Fig. 10a). Pathway enrichment identified two major functions: (1)

inflammation, including complement and coagulation, innate-immune system, lipoprotein remodelling and cytokine and TNF/NF-κB signalling (*Ccl2*, *Ccl7*, *Ccl8* and *Isg15*); and (2) fibrosis, including matrix organization and collagen metabolism, and TGFβ signalling (*Col3a1*, *Col6a2* and *Timp2*) (Fig. 5a and Extended Data Fig. 10b,c). Some of these genes were commonly identified in scRNA-seq (Fig. 1g). IGF regulation by IGF-binding proteins (*Igfbp4*, *Igfbp6*, *Igfbp7* and *Igf1*) was also present (Fig. 5a and Extended Data Fig. 10b,c). Thus, the major SASP features corresponded with the universal hallmarks of senescent cells in vivo (Fig. 4a).

Fig. 5: Senescent cells create an aged-like microenvironment in young regenerative niches through pro-inflammatory and pro-fibrotic factor secretion.



a, SASP-related gene set enriched clusters from Sen SCs, FAPs and MCs of young mice at 3 d.p.i. (FDR < 0.05). The grey edges indicate gene overlap. Differentially upregulated genes (FDR < 0.05) were considered to be SASP genes when overexpressed in Sen versus their NSen populations. **b**, Comparison of enrichments for differential RNA expression in tissues from aged mice, rats, humans, killifish and senescent populations from young 3 d.p.i. muscles. Cbm, cerebellum. **c**, The overlap between secreted proteins in cells of young and aged muscle. **d**, Chord diagram showing transcription factors that regulate SASP genes and their categories in Sen versus NSen. Chord width is proportional to the significance ($-\log_{10}$[FDR]) of GO molecular function (GO:MF) cluster enrichment. ECM, extracellular matrix. **e**, SPiDER$^+$ and SPiDER$^-$ cells from 3 d.p.i. young muscle were stained with Dil and transplanted into the preinjured TA of recipient mice for 4 days. $n = 4$ mice per group. Quantification of SA-β-gal$^+$ and CD11b$^+$ cells, and Sirius Red staining. **f**, As described in **e**, but senescent and non-senescent C2C12 cells were transplanted into young p16-3MR mice for 5 days. Images and quantification of luciferase-activity. $n = 4$ mice (non-senescent) and $n = 5$ mice (senescent). **g**, As described in **e**, but quantification of γH2AX in Pax7$^+$ cells ($n = 109$ cells (SPiDER$^-$) and $n = 100$ cells (SPiDER$^+$)). **h**, CTX-injured p16-3MR mice were treated daily with GCV or vehicle until 4 d.p.i. ($n = 5$ TA muscles from 3 mice (vehicle) and $n = 6$ TA muscles from 3 mice (GCV)). Left, images and quantification of EdU or Pax7 staining. Right, BrdU incorporation in SCs in vitro. At 3 d.p.i., SCs were sorted and cultured for 3 days. $n = 4$ mice per group. **i**, SPiDER$^-$ SCs were isolated from 3 d.p.i. muscles, and cultured for 3 days in Transwells with total SPiDER$^+$ or SPiDER$^-$ cells or culture medium, and analysed for BrdU incorporation. $n = 3$ mice per group. **j**, EDL muscles from WT or p16-3MR-donor mice were transplanted into WT or p16-3MR recipient mice (or vice versa). The recipient mice were treated daily with GCV, and regeneration was analysed at 7 d.p.i. The CSA of MYH3$^+$ fibres ($n = 8$ (WT/WT), $n = 6$ (p16-3MR/WT) and $n = 7$ (other groups) mice) and SA-β-gal$^+$ cells ($n = 7$ mice per group). **k**, SPiDER$^+$ and SPiDER$^-$ SCs, FAPs or MCs ($n = 5$ mice (SCs) and $n = 4$ mice (FAPs and MCs)) transplanted as in **e**. The CSA of MYH3$^+$ fibres. For **e**–**k**, data are mean ± s.e.m. *P* values were calculated using Tukey’s tests (**i** and **j**) and Mann–Whitney *U*-tests (**e**–**h** and **k**).

Source data

We next compared the transcriptomes of ageing tissues from various species (including humans) with those of Sen cells from injured young muscles, and found an increase in inflammatory pathways (interferon, complement and cytokine and TNF/NF-κB signalling) (Fig. 5b), consistent with previously published datasets[36,37,38]. A secreted-protein array-based assay confirmed the secretion of inflammatory/matrix-remodelling proteins in sorted Sen cells (from individual cell types or in combination) from young muscle (Fig. 5c and Extended Data Fig. 10d,e). Many of these proteins were also secreted by Sen cells from injured aged muscle (CCL2, IGFBP6, CD40, IL13, CXCL1, IL1A, CXCL11 and MMP2) (Fig. 5c and Extended Data Fig. 10d,e). In whole-tissue analysis, many secreted SASP proteins were commonly upregulated in injured young muscles and in non-injured aged muscles compared with non-injured young muscles (Extended Data Fig. 10f), indicating a shared inflammatory secretome in injured young tissue and in basal ageing conditions. Thus, the SASP of Sen cells, transiently present in injured young muscles, mimics aged-like inflammageing, which is exacerbated in injured aged muscle.

Functional profiling of SASP transcription factors confirmed the association of NF-κB, STAT1/3 and SMAD3/4 with the identified inflammatory and matrix-related SASP genes (Fig. 5d). To assess the role of NF-κB and SMAD3 in the inflammatory SASP in regenerating muscles, we treated young mice with either NF-κB (bortezomib) or SMAD3 (SIS3) inhibitors after injury. Both bortezomib and SIS3 diminished the expression of pro-inflammatory and pro-fibrotic SASPs in sorted SPiDER$^+$ cells, on the basis of analyses using a secreted-protein array and gene-transcription assays (Extended Data Fig. 11a–d), suggesting that NF-κB and SMAD3 link injury-induced senescence and inflammageing in vivo.

SASP reduces muscle stem cell expansion

To assess how the SASP affects muscle regeneration (and especially muscle SCs), we reconstructed ligand–receptor interactions between Sen cells (producing SASP ligands) and NSen SCs (with cognate receptors) using a modified FunRes algorithm[39]; this revealed predominantly inflammatory

interactions (Extended Data Fig. 11e). A signalling pathway impact analysis (SPIA)[40] of transcription factors involved in these interactions revealed that the SASP produced by all Sen cells activated downstream signalling pathways in receiving NSen SCs (senescence, apoptosis and inflammatory responses) and inhibited cell-cycle and proliferative pathways (MAPK and AKT signalling) (Extended Data Fig. 11f). Thus, SASP components might provoke either proliferative arrest or paracrine senescence in NSen SCs. To test these predictions, we transplanted sorted Sen (SPiDER$^+$) and NSen (SPiDER$^-$) fractions, labelled with Dil, into preinjured muscles of recipient mice. In contrast to transplanted Dil-labelled NSen cells, transplanted Dil-labelled Sen cells increased the number of senescent cells in the host tissue, induced inflammatory-cell recruitment and fibrosis, and delayed regeneration (Figs. 2d and 5e). This paracrine senescence induction was confirmed by transplantation of ex vivo induced senescent (versus non-senescent) WT cells into preinjured muscles of p16-3MR mice (Fig. 5f). Notably, transplantation of sorted SPiDER$^+$ and SPiDER$^-$ into non-injured muscles was sufficient to induce DNA damage in endogenous SCs (Fig. 5g). By contrast, GCV-treated p16-3MR mice showed higher numbers of proliferating and total SCs within the regenerative-muscle niche (Fig. 5h; Extended Data Fig. 11g). Consistently, SCs sorted from GCV-treated p16-3MR mice had a higher proliferation ability ex vivo compared with SCs from vehicle-treated mice (Fig. 5h) and, in Transwell assays, Sen cells reduced the proliferation of NSen SCs (Fig. 5i and Extended Data Fig. 11h). Thus, senescent cells restrained muscle regeneration through paracrine pro-inflammatory and pro-fibrotic SASP functions that blunted SC proliferation.

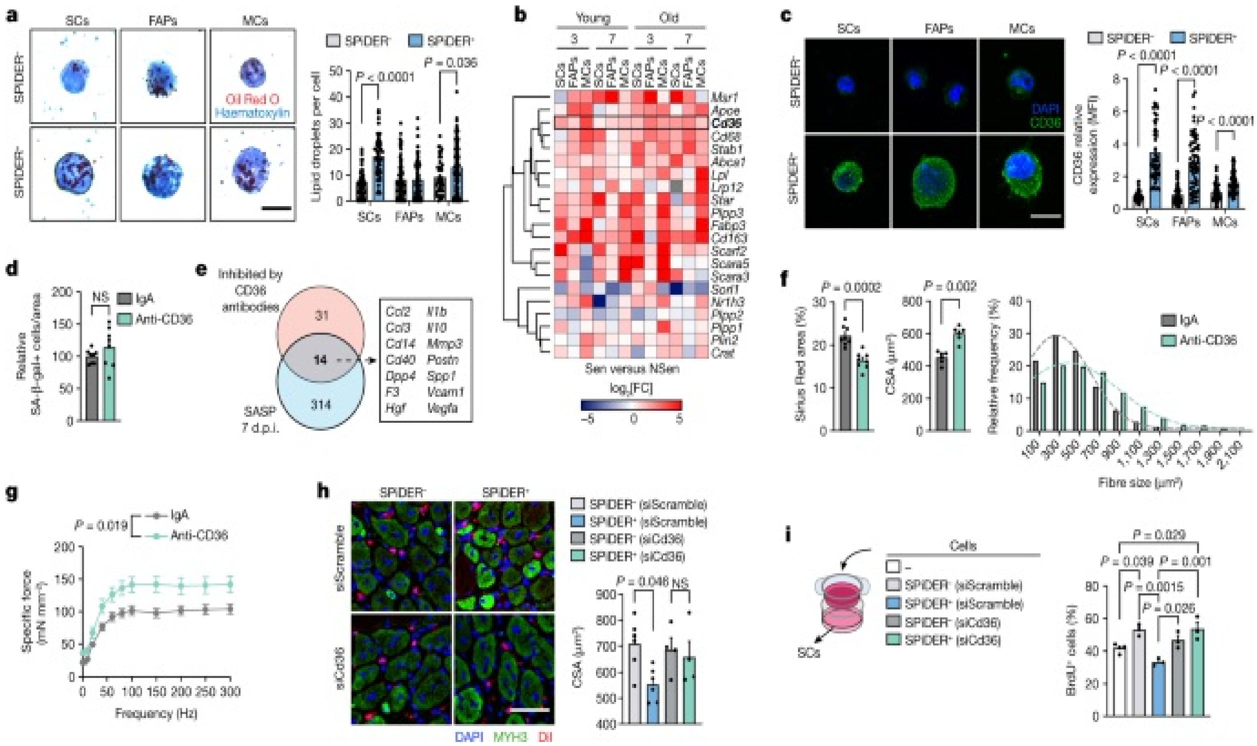
To examine the relative contribution to regeneration of senescent cells derived either from tissue-resident cells or from the blood, we used a whole-muscle grafting model[5], in which the extensor digitorum longus (EDL) muscle from one mouse was grafted onto the tibialis anterior (TA) muscle of a recipient mouse. In this model, the transplanted EDL undergoes de novo myogenesis at the expense of its own SCs, while recruited bone-marrow-derived cells come from the host. EDL grafting combined with daily GCV-mediated senescent-cell depletion revealed that p16-3MR EDL grafts in WT host mice, or WT EDL grafts in p16-3MR hosts, had larger regenerating myofibres compared with WT EDL grafts in WT mice (Fig. 5j and Extended Data Fig. 11i), confirming that senescent cells from EDL-resident SCs and

FAPs, and from blood cells, hampered muscle regeneration. Similar detrimental effects were seen after transplantation of equal numbers of SPiDER$^+$ SCs, FAPs or MCs, separately or combined (Figs. 2d and 5k), supporting that the SASP was not beneficial for regeneration even if acting transiently.

CD36 regulates SASP production

Given the heterogeneity of SASPs in vivo, we searched for a broad SASP-targeting approach. Lipid transport, which is tightly associated with inflammatory responses[41], was consistently included in the inflammatory hallmark of Sen cells (Fig. 4a) and the SASP (Fig. 5a and Extended Data Fig. 10b) in all conditions. Sen cells had more lipid droplets compared with NSen cells (Fig. 6a). Numerous lipid metabolism and lipid-transport genes, including *Fabp3*, *Apoe*, *Star*, *Lpl*, *Cd68* and *Cd36*, were upregulated in all Sen cells (Fig. 6b). As lipid uptake and CD36 are related to SASP in vitro[42,43], CD36 might also be related to the SASP in vivo. FunRes generated a CD36 signalling network that predicted downstream activation of NF-κB and other inflammation/stress-related pathways, and downstream SASP components, such as *Il6*, *Tgfb1*, *Mmp3*, *Igfbp5*, *Ccl2* and *Cxcl10* (Extended Data Fig. 12a), suggesting that CD36 might regulate the in vivo senescence secretory program, affecting regeneration.

Fig. 6: CD36 neutralization improves muscle regeneration through a senomorphic action.



a, SPiDER$^+$ and SPiDER$^-$ populations from 3 d.p.i. muscle stained with Oil Red O and haematoxylin. Images and lipid-droplet quantification. *n* = 90 SCs (NSen), 51 SCs (Sen), 89 FAPs (NSen), 90 FAPs (Sen), 45 MCs (NSen) and 94 MCs (Sen). **b**, Heat map showing lipid-transport-related genes that are differentially expressed in at least 3 out of 12 comparisons between Sen and NSen cells. The colour indicates the log_2-transformed fold change in expression. **c**, Images and CD36 quantification in SPiDER$^+$ and SPiDER$^-$ populations from 3 d.p.i. muscles. *n* = 44 SCs (NSen), 50 SCs (Sen), 56 FAPs (NSen), 62 FAPs (Sen), 55 MCs (NSen) and 56 MCs (Sen). **d**, SA-β-gal$^+$ cells in the injured TA area of aged mice (treated with anti-CD36 antibodies or control IgA). *n* = 8 TA from 4 mice for both groups. **e**, The overlap between SASP-upregulated genes in 7 d.p.i. senescent populations of old mice, and those reduced by anti-CD36 treatment. **f**, The CSA and frequency distribution of MYH3$^+$ fibre size (*n* = 6 TA from 4 mice for both groups) and Sirius Red staining (*n* = 8 TA from 4 mice for both groups) of 7 d.p.i. TA from old mice treated with anti-CD36-antibody antibodies or control IgA. **g**, Injured EDL of old mice treated with anti-CD36 antibodies or control IgA. (*n* = 6 EDL from 3 mice (IgA) and *n* = 5 EDL from 4 mice (anti-CD36)). Force–frequency curves are shown. **h**, SPiDER$^+$ and SPiDER$^-$ cells from 3 d.p.i. muscles were transfected with siCd36 or siScramble, stained with Dil

and transplanted into the preinjured recipient TA. Images and the CSA of MYH3$^+$ fibres 4 days after transplantation. $n = 6$ (siScramble-treated SPiDER$^+$ and SPiDER$^-$ cells) and $n = 4$ (siCd36-treated SPiDER$^+$ and SPiDER$^-$ cells) mice. **i**, SCs (from 3 d.p.i. muscles) were cultured for 3 days in Transwells with senescent or non-senescent C2C12 cells that were previously treated with siCd36 or siScramble, or without cells. BrdU incorporation is shown. $n = 4$ (empty Transwell) and $n = 3$ (other groups) mice. Scale bars, 10 µm (**a** and **c**) and 50 µm (**h**). For **a**, **c**, **d** and **f**–**i**, data are mean ± s.e.m. *P* values were calculated using Mann–Whitney *U*-tests (**a**, **c** and **f**), mixed-effects analysis (**g**) and Tukey's tests (**h** and **i**); NS, not significant.

Source data

All three Sen cell types had higher CD36 protein expression in injured muscles (Fig. 6c). *Cd36* expression was also upregulated in etoposide-induced senescent cells in vitro (Extended Data Fig. 12b). We next analysed injured muscles from young or old mice, treated for 4 days (starting at 3 d.p.i.) with an anti-CD36 neutralizing antibody (at two distinct doses) or a control IgA antibody. CD36 blockade did not affect the number of senescent cells (Fig. 6d and Extended Data Fig. 12c), but reduced several SASP proteins (Extended Data Fig. 12d), and many coincided with SASP genes in Sen SCs, FAPs and MCs encoding chemokines, cytokines and matrix metalloproteinases (*Ccl2*, *Ccl3*, *Il1b*, *Il10* and *Mmp3*) (Fig. 6e). Moreover, the whole injured-muscle secretome showed lower SASP protein levels in both GCV-treated and anti-CD36-antibody-treated p16-3MR mice (Extended Data Fig. 12e). Several Sen-cell-secreted inflammatory SASPs (*Ccl2*, *Ccl4* and *Cxcl10*) that induced downstream signalling in NSen SCs in the ligand–receptor interactive network were predicted to negatively affect SC functions and regeneration (Extended Data Fig. 11e,f), and some appeared as CD36-regulated SASPs (Extended Data Fig. 12a). CD36 blockade improved regeneration in both young and old muscles (Fig. 6f and Extended Data Fig. 12c,f) while reducing inflammation (Extended Data Fig. 12g) and fibrosis (Fig. 6f and Extended Data Fig. 12c,f), and these muscles showed increased force (Fig. 6g and Extended Data Fig. 12h).

We next silenced *Cd36* in sorted Sen cells using *Cd36*-specific short

interfering RNA (siCd36), with scrambled siRNA (siScramble) used as a control. After transplantation into injured muscle, siScramble-treated Sen cells delayed regeneration, whereas *Cd36*-silenced Sen cells had no negative effects (Fig. 6h). Consistent with this result, the SASPs produced by sorted Sen cells reduced SC proliferation in co-culture Transwell assays (Fig. 5i), but this effect was not observed when *Cd36* was silenced in Sen cells before co-culturing with SCs (Fig. 6i); thus, CD36 was crucial for the paracrine effects of Sen cells on muscle regeneration by regulating SASP production.

Discussion

Proper reconstruction of injured tissue requires timely interaction among diverse cell types within the regenerative niche. Senescent cells affect tissue repair processes, but the mechanisms are largely unclear. One reason has been the technical limitations imposed by both the scarcity and the heterogeneity of senescent cells, and a lack of 'universal' senescence markers[3]. By establishing a sorting protocol that enriches senescent cells of distinct types (confirmed in vivo using multi-marker microscopy), we have now identified that senescent cells are integral components of the skeletal muscle niche and have demonstrated that they repress regeneration in response to distinct types of injuries in young mice, and even more strongly in aged mice.

Here we showed that senescent cells were practically absent (or in very low numbers) in unperturbed muscle tissues, even in old age, but emerged after injury. In injured muscle, senescent cells modified their normal niche counterparts by creating an aged-like inflamed microenvironment that hijacked their proliferative programs and blunted regeneration. Reducing the number of senescent cells (and therefore their inflammatory secretome) resumed stem-cell proliferation and enhanced muscle regeneration. Thus, efficient tissue regeneration involves not only constructive cellular crosstalks but also unexpectedly repressive interactions.

An important question is how senescent cells arise after tissue damage. We provide evidence that, after injury, a subset of niche cells accumulates ROS and DNA damage beyond a threshold, which leads to senescence and regenerative failure. Thus, proliferating cells in injured tissues actively

repress their senescence program, thereby preserving cell integrity and fitness. This is particularly critical in old age as long-lasting cells accumulate damage (including telomere damage) over a lifetime and are therefore primed for senescence. Furthermore, the functionally weakened immune system in old age[44] does not clear senescent cells, which may account for their persistence in damaged aged muscle; this exacerbates tissue regenerative impairment.

Senescence has been molecularly characterized mainly in vitro[10,45,46]. Through FACS-based enrichment of SA-β-gal-positive cells from injured muscle, followed by scRNA-seq, we generated a map of senescent cells in vivo, revealing three major types of senescent cells in regenerating muscle (MCs, FAPs and SCs), as well as other smaller populations. The successful enrichment of bona fide senescent cells in our method relied on restricting the FACS-isolation protocol exclusively to the highest SPiDER-β-gal expressing cells. Yet, some classical markers of senescence in vitro (such as CDKN2A, $p21^{CIP1}$, IL-6 and IL-1) were observed in some, but not in all, senescent cells in vivo. A possible explanation is that the senescent state can be influenced in vivo by its trigger, the cell that senesces, its environment and its temporal resolution by the immune system, resulting in a wide spectrum of phenotypic features. We confirmed that MCs, FAPs and SCs (in this order of abundance) are the principal senescent cell types in regenerating muscle. Despite the variability in gene expression within senescent cells, we revealed that two major conserved universal hallmarks define senescent cells: inflammation and fibrosis. We also determined that, depending on their lineage of origin, senescent cells have unique features that are preserved throughout life, indicating that they share universal programs while maintaining identity traits, and adapt to the idiosyncrasies of their origin and age. Thus, our study provided a better definition of senescence in vivo: despite being a state of irreversible arrest, senescence encompasses features of molecular diversity and dynamism (sensitive to the pass of time and mode of injury) as well as conserved hallmarks.

Our pathway analyses revealed that inflammation is also broadly upregulated in senescent cells, which is consistent with the concept of inflammageing[4]. We detected upregulation of complement, interferon and TNF/NF-κB pathways in senescent cells, which are robustly activated with ageing across

vertebrate tissues, including humans[36,37,38]. Notably, interferon signalling has been linked to viral infection responses and mitochondrial stress[47,48], and to the reactivation of transposable elements, which is also linked to senescence[49]. Transcriptomic and chromatin-accessibility analyses in senescent cells in vivo revealed an association between the transcription factors NF-κB, SMAD, IRF1/3 and C/EBPβ, and the induction of inflammatory and matrix components and interferon-response genes, accounting for the deleterious SASP effects on niche cells. In this context, NF-κB and SMAD3 (signalling participants in inflammageing[4,38]) appeared to be causally involved in the pro-inflammatory/pro-fibrotic SASP production by senescent cells in young injured muscles.

Thus, our studies demonstrate that senescent cells are a decisive factor in tissue regeneration throughout life. The transformation of subsets of SCs and niche cells into permanently arrested senescent cells not only reduces the progeny available for regeneration but also, through the pro-inflammatory SASP, confers young tissue with an aged-like inflamed niche, therefore mirroring the negative effects of inflammageing. Mechanistically, the SASP induces senescence in surrounding healthy cells, which further spreads senescence and lowers stem-cell proliferation. Consistent with this, a reduction in the number of senescent cells improved regeneration as a result of attenuating inflammatory and fibrotic SASPs. Finally, we identified CD36 (a scavenger receptor that is related to lipid metabolism and inflammatory function) as a cell receptor acting as a senomorphic in vivo. Together, these results challenge the prevailing idea that cellular senescence is always beneficial when it occurs transiently in young tissues, and deleterious only when it occurs chronically during ageing or in diseased states. Notably, outside the tissue regeneration context (this study), both positive and negative roles for senescent cells have been proposed in muscle, such as in exercised muscle, during Yamanaka-factor-induced muscle reprogramming and after irradiation[17,22,50], therefore exemplifying the complexity of the functions of senescent cells in distinct in vivo contexts.

Through the technical advancement in senescent-cell separation, our findings provide a transcriptome atlas of specific senescent-cell types in vivo (and a common SASP signature), and a conceptual explanation for the emergence, causes, definition, dynamics and consequences of senescence in vivo. We

recognize limitations in our study associated to: (1) the use of distinct senescent-cell ablation methods (genetic and pharmacologic) and (2) the presence of a small percentage of non-senescent cells in the SPiDER^{+}-sorted fraction. However, our findings provide evidence that senescence is a more complex array of states than previously anticipated. As senescent cells also emerge in damaged human muscle, these findings have implications for regenerative medicine, including in sarcopenia.

Methods

Animal models

C57Bl/6 (WT), p16-3MR (donated by J. Campisi)[18], dystrophic mdx (DBA/2-background) and mdx/p16-3MR (dystrophic mdx mice crossed with p16-3MR mice) were bred and aged at the animal facility of the Barcelona Biomedical Research Park (PRBB), housed in standard cages under 12 h–12 h light–dark cycles and fed ad libitum with a standard chow diet. All of the experiments followed the principle of the 'three Rs'—replacement, reduction and refinement according to Directive 63/2010 and its implementation in the Member States. All of the procedures had authorization from the PRBB Animal Research Ethics Committee (PRBB-CEEA) and the local government (Generalitat de Catalunya) and were conducted according to the European Directive 2010/63/EU and Spanish regulations RD 53/2013. Both male and female mice were used in each experiment unless stated otherwise. Live colonies were maintained and genotyped according to the Jackson Laboratories' guidelines and protocols. The mice were housed together, their health was monitored daily for sickness symptoms (not age-related weight loss and so on) and they were euthanized immediately at the clinical end point when recommended by veterinary and biological service staff members. Mice were randomly allocated to experimental or treatment groups. No blinding was used. No statistical methods were used to predetermine the sample size. For PCR genotyping, the following primers were used: p16-3MR-1, 5′-AACGCAAACGCATGATCACTG-3′; and p16-3MR-2, 5′-TCAGGGATGATGCATCTAGC-3′. Positive mice show a band at 202 bp.

Human biopsies

Human muscle biopsy samples from the vastus lateralis muscle of patients undergoing surgery were obtained from the biobank of the EU/FP7 Myoage Consortium, as previously reported[5,6]. Ethical approval was received from the local ethics committees at each of the five research centres of the Consortium. All of the participants provided written informed consent and were medically screened before participation. The biobanked muscle tissue had been directly frozen in melting isopentane and stored at −80 °C until analysed. Damaged areas were identified by morphological criteria on the basis of the presence of infiltrating mononuclear cells. Data are from female patients aged 69, 82, 80, 89 or 85 years old; the average age was 81 ± 7.5 years.

In vivo treatments

Quercetin (USP, 1592409; 50 mg kg^{-1}) and dasatinib (LC Laboratories, D-3307; 5 mg kg^{-1}) were administered orally (by gavage). Control mice were administered with an equal volume of vehicle (10% ethanol, 30% polyethilenglicol and 60% phosal). GCV (Sigma-Aldrich, G2536-100MG; 25 mg kg^{-1}) was injected intraperitoneally (i.p.). Anti-CD36 antibodies (Cayman Chemical, 10009893; 10 µg or 20 µg in young mice and 20 µg in old mice) diluted in PBS was administered via i.p., and control mice received an equal dose of IgA control antibodies (Southern Biotech/Bionova, 0106-14). Treatments with GCV, senolytics and CD36 were administered daily for 4–7 consecutive days as indicated in the figure legends. NAC (Sigma-Aldrich, A9165; 0.01 g ml^{-1}) was added into drinking water (exchanged every 3 days) 1 week before muscle injury and was prolonged until euthanasia. SIS3 (Sigma-Aldrich, S0447-5MG; 10 mg kg^{-1}) in PBS and bortezomib (Teva, TH000345/LF2.3; 0.5 mg kg^{-1}) in 10% DMSO diluted in PBS were administered i.p. (from 1 to 5 d.p.i.). For long-term treatments, 3-month-old mdx and mdx/p16-3MR mice were administered with D+Q or GCV, respectively, twice a week for 2 months.

Muscle regeneration

Mice were anaesthetized with ketamine–xylazine (80 and 10 mg kg^{-1}

respectively; i.p.) or isoflurane. Regeneration of skeletal muscle was induced by intramuscular injection of CTX (Latoxan, L8102; 10 µM) as described previously[51]. At the indicated times after injury, the mice were euthanized and the muscles were dissected, frozen in liquid-nitrogen-cooled isopentane, and stored at −80 °C until analysis.

Heterografting

Heterografting experiments were performed according to the protocol described previously[52]. In brief, the EDL muscle was removed from the anatomical bed of either p16-3MR or WT mice and was transplanted onto the surface of the TA muscle of the p16-3MR or WT recipient mouse or vice versa. Muscle grafts were collected on day 7 after transplantation.

Muscle force measurement

Ex vivo force measurements of EDL muscles were assessed as previously described[53] using and 300B apparatus (Aurora Scientific). Force was normalized per muscle area, determined by dividing the muscle mass by the product of length and muscle density of (1.06 mg mm^{-3}), to calculate the specific force (mN mm^{-2}).

p16-3MR *Renilla* luciferase reporter assay

In vivo, *Renilla* luciferase activity was measured in the TA, quadriceps and gastrocnemius muscles of p16-3MR mice. Anaesthetized mice were injected intramuscularly with coelenterazine H (PerkinElmer, 760506) and luciferase activity was immediately measured using the IVIS Lumina III (PerkinElmer) system. In vitro, *Renilla* luciferase activity was measured from the cryopreserved diaphragm and TA muscles using the Dual-Luciferase Reporter Assay Kit (Promega, E1910). The signal was measured using the luminometer Centro LB 960 (Berthold Technologies) and values were normalized to the total protein extracted measured using Bradford method (Protein Assay, Bio-Rad, 500–0006), and the damaged area was measured after haematoxylin and eosin (H&E) staining.

Cell isolation by flow cytometry

Muscles were mechanically disaggregated and incubated in Dulbecco's modified Eagle's medium (DMEM) containing liberase (Roche, 177246) and dispase (Gibco, 17105-041) at 37 °C with agitation for 1–2 h. When required, SPiDER-β-gal reagent (Dojindo, SG02; 1 μM) was added during the second hour. The supernatant was then filtered and cells were incubated in lysis buffer (BD Pharm Lyse, 555899) for 10 min on ice, resuspended in PBS with 2.5% fetal bovine serum (FBS) and counted. BV711-conjugated anti-CD45 (BD, 563709; 1:200), APC-Cy7-conjugated anti-F4/80 (BioLegend, 123118; 1:200), PE-conjugated anti-α7-integrin (Ablab, AB10STMW215; 1:200), APC-conjugated anti-CD31 (eBioscience, 17-0311-82; 1:200) and PE-Cy7-conjugated anti-SCA1 (BioLegend, 108114; 1:200) antibodies were used to isolate MCs ($CD45^{+}F4/80^{+}$), SCs (α7-$integrin^{+}CD45^{−}F4/80^{−}CD31^{−}$) and FAPs ($SCA1^{+}CD45^{−}F4/80^{−}$α7-$integrin^{−}CD31^{−}$). PE-Cy7-conjugated anti-CD45 antibodies (BioLegend, 103114) were used to isolate CD45-positive and CD45-negative populations (Extended Data Fig. 1h). SPiDER-β-gal (SPiDER) was used to isolate senescent cells ($SPiDER^{+}$) from non-senescent cells ($SPiDER^{−}$) of each cell type (Extended Data Figs. 1h and 4a) (Supplementary Table 1). Cells were sorted using the FACS Aria II (BD) system. Cell lineage was confirmed by specific-cell marker expression (Extended Data Fig. 4b–d). Isolated cells were used either for RNA extraction, cell cultures, engraftments, proliferation assays or plated onto glass slides (Thermo Fisher Scientific, 177402) for immunostaining and SA-β-gal analysis.

To isolate ROS^{high} and ROS^{low} populations, the digested muscle was stained with CellRox Green reagent (Invitrogen, C10444; 5 μM) according to the manufacturer's protocol and PE-Cy7-conjugated anti-CD45 (BioLegend, 103114; 1:200), PE-Cy7-conjugated anti-CD31 (BioLegend, 102418; 1:200), PE-conjugated anti-α7-integrin (Ablab, AB10STMW215; 1:200) and APC-conjugated anti-SCA1 (BioLegend, 108111; 1:200) antibodies to separate SCs (α7-$integrin^{+}CD45^{−}CD31^{−}$) and FAPs ($SCA1^{+}CD45^{−}$α7-integrin $^{−}CD31^{−}$) (Supplementary Table 1). $CellRox^{high}$ and $CellRox^{low}$ cells were sorted using the FACS Aria II (BD) system. Isolated cells were used for cell cultures and proliferation assays. The acquisition was performed using the

BD FACS Diva software.

Senescent cell transplantation

Cells transplants were performed as described previously[5], following an adapted protocol[54]. FACS-isolated SPiDER$^+$ and SPiDER$^−$ cells were collected, resuspended in 20% FBS DMEM medium, labelled with Vybrant Dil Cell Labelling solution (Invitrogen, V22889) according to manufacturer instructions and injected into the TA muscles of recipient mice that were either uninjured or previously injured using the freeze crush method 2 days before[55]. The cell-type proportions of MCs, SCs and FAPs were controlled in the transplanted SPiDER$^+$ and SPiDER$^−$ populations. Each TA muscle was engrafted with 10,000 cells, except when each senescent cell type was transplanted separately (Fig. 5k), where 5,000 cells were engrafted. Engrafted muscles were collected and processed for muscle histology 4 days after cell transplantation.

RNA interference

Freshly sorted cells or C2C12 cells (ATCC, CRL-1772) were transfected with siRNA targeting *Cd36* (On-Target plus SmartPool, Dharmacon, L-062017-00-0005; 5 nM) or unrelated sequence as control (On-Target plus non-targeting siRNA Pool, Dharmacon, D-001810-10-05; 5 nM) using the DharmaFect protocol (Dharmacon, T-2003-02). Target sequences for *Cd36* siRNA were as follows: 5′-CCACAUAUCUACCAAAAUU-3′, 5′-GAAAGGAUAACAUAAGCAA-3′, 5’-AUACAGAGUUCGUUAUCUA-3’, 5’-GGAUUGGAGUGGUGAUGUU-3’. Freshly sorted cells after incubation with siRNAs for 3 hours were washed and engrafted.

Cytokine array

Cytokine antibody arrays (R&D Systems, ARY028; Abcam, ab193659) were used according to the manufacturer’s protocol. For cells, freshly sorted cells were cultured for 24 h in serum-free DMEM. Cell culture supernatants were collected, centrifuged and incubated with the membranes precoated with

captured antibodies. For tissue interstitial fluid, skeletal muscles of mice were dissected and slowly injected with a PBS solution with a Complete Mini EDTA-free protease inhibitor cocktail (Roche, 11836170001). The PBS exudate was then recovered centrifuged and incubated with the membranes precoated with captured antibodies. The membranes were then incubated with detection antibodies, streptavidin–HRP and Chemi Reagent Mix. The immunoblot images were captured and visualized using the ChemiDoc MP Imaging System (Bio-Rad) and the intensity of each spot in the captured images was analysed using the publicly available ImageJ software.

Proliferation assays

To assess proliferation in vivo, muscles were injured by local CTX injection, and the mice were administered with ethynyl-labelled deoxyuridine (EdU, Invitrogen, A10044; 25.5 mg kg^{-1}; i.p.) 2 h before euthanasia at 4 d.p.i. The muscles were collected and processed for immunofluorescence staining in tissue slides or cell isolation by FACS. EdU-labelled cells were detected using the Click-iT EdU Imaging Kit (Invitrogen, C10086). EdU-positive cells were quantified as the percentage of the total number of cells analysed. In vitro proliferation was quantified on freshly sorted SCs, seeded in 20% FBS Ham's F10 medium supplemented with b-FGF (Peprotech, 100-18B-250UG; 2.5 ng ml^{-1}) in collagen-coated plates. After 3 days of culture, SCs were pulse-labelled with bromodeoxyuridine (BrdU, Sigma-Aldrich, B9285-1G; 1.5 μg ml^{-1}) for 1 h. BrdU-labelled cells were detected by immunostaining using rat anti-BrdU antibodies (Abcam, AB6326, 1:500) and a specific secondary biotinylated donkey anti-rat antibody (Jackson Immunoresearch, 712-066-150, 1:250). Antibody binding was visualized using Vectastain Elite ABC reagent (Vector Laboratories, PK-6100) and 3,3′-diaminobenzidine. BrdU-positive cells were quantified as the percentage of the total number of cells analysed.

Transwell assay

SCs were freshly isolated from regenerating muscle tissue at 3 d.p.i. and plated onto 24-well plates (Falcon, 353047) in 20% FBS DMEM supplemented with b-FGF. Subsequently, medium or freshly sorted $SPiDER^{+}$

and SPiDER^{-} cell populations (Fig. 5i) or etoposide-induced senescent C2C12 cells (Fig. 6i) were seeded on a 0.4-μm-pore-size cell culture insert (Falcon, 353495) using the same medium. After 3 days of culture, a proliferation assay was performed on SCs with BrdU labelling as described above.

In vitro treatments

ROS^{high} and ROS^{low} SCs and FAPs were freshly isolated from regenerating muscle at 24 h after injury, seeded and cultured in the presence of NAC (10 mM) or vehicle for 3 days. After the treatment, cells were fixed and further processed for staining. C2C12 cells maintained in 10% FBS DMEM were treated with etoposide (Sigma-Aldrich, E1383, 1 μM) for 5 days to induce senescence and were collected for RNA extraction and RT–qPCR. Cells were stained with a β-galactosidase staining kit (as described below) to confirm their senescent state.

Cell staining

SA-β-galactosidase (SA-β-gal) activity was detected in freshly sorted cells and cell cultures using the senescence β-galactosidase staining kit (Cell signalling, 9860) according to the manufacturer's instructions. Lipid droplets were stained with Oil Red O (Sigma-Aldrich, O0625) according to manufacturer instructions. ROS levels were measured by immunofluorescence using CellRox Green reagent (Invitrogen, C10444; 5 μM) according to the instructions. TUNEL assays were performed using the In Situ Cell Death Detection Kit, Fluorescein (Roche, 11684795910), cells treated with DNase were used as a positive control of the staining according to the manufacturer's description.

Muscle histology, immunofluorescence and immuno-FISH

Muscles were embedded in OCT solution (TissueTek, 4583), frozen in isopentane cooled with liquid nitrogen and stored at −80 °C until analysis.

Muscle cryosections (thickness, 10 μm) were collected and stained for SA-β-gal (AppliChem, A1007,0001), H&E (Sigma-Aldrich, HHS80 and 45235), MYH3 (DSHB, F1.652), Sirius Red (Sigma-Aldrich, 365548) or used for immunofluorescence (Supplementary Table 1). The CSA of H&E- and MYH3-antibody-stained sections, the percentage of muscle area positive for Sirius Red staining and the number of SA-β-gal^{+} cells were quantified using Image J. Double immunofluorescence was performed by the sequential addition of each primary and secondary antibody using positive and negative controls. The sections were air-dried, fixed, washed on PBS and incubated with primary antibodies according to the standard protocol after blocking with a high-protein-containing solution in PBS for 1 h at room temperature. Subsequently, the slides were washed with PBS and incubated with the appropriate secondary antibodies and labelling dyes. Telomere immuno-FISH was performed after γH2AX immunofluorescence staining with telomeric PNA probe (Panagene, F1002-5) as described previously[56].

Digital image acquisition

Digital images were acquired using an upright DMR6000B microscope (Leica) with a DFC550 camera for immunohistochemical colour pictures; a Thunder imager 3D live-cell microscope (Leica Microsystems) with hardware autofocus control and a Leica DFC9000 GTC sCMOS camera, using HC PL FLUOTAR ×10/0.32 PH1 ∞/0.17/ON257C and HC PL FLUOTAR ×20/0.4 CORR PH1 ∞/0-2/ON25/C objectives; a Zeiss Cell Observer HS with a ×20 and x40 air objective and a Zeiss AxioCam MrX camera; and a Leica SP5 confocal laser-scanning microscope with HCX PL Fluotar ×40/0.75 and ×63/0.75 objectives. The different fluorophores (three or four) were excited using the 405, 488, 568 and 633 nm excitation lines. The acquisition was performed using the Leica Application (v.3.0) or LAS X (v.1.0) software (Leica) or Zeiss LSM software Zen 2 Blue.

RNA isolation and RT–qPCR

Total RNA was isolated from snap-frozen muscles using the miRNAeasy Mini Kit (Qiagen, 1038703). PicoPure (Thermo Fisher Scientific, KIT0204) was used for RNA isolation from sorted cells. For RT–qPCR experiments,

DNase digestion of 10 mg of RNA was performed using 2 U DNase (Qiagen, 1010395). cDNA was synthesized from total RNA using SuperScript III Reverse Transcriptase (Invitrogen, 18080-044). For gene expression analysis in freshly sorted SCs, FAPs and MCs, cDNA was pre-amplified using the SsoAdvanced PreAmp Supermix (Bio-Rad, 172-5160) according to the manufacturer's instructions. qPCR reactions were performed as described previously[57]. Reactions were run in triplicate, and automatically detected threshold cycle values were compared between samples. Transcripts of the *Rpl7* housekeeping gene were used as the endogenous control, with each unknown sample normalized to *Rpl7* content (a list of the primers used in this study is provided in Supplementary Table 2).

RNA-seq sample and library preparation

Sequencing libraries were prepared directly from the lysed cells, without a previous RNA-extraction step. RNA reverse transcription and cDNA amplification were performed using the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing from Clontech Takara. The Illumina Nextera XT kit was used for preparing the libraries from the amplified cDNA. Libraries were sequenced using the Illumina HiSeq 2500 sequencer (51 bp read length, single-end, around 20 million reads).

Bulk RNA-seq data preprocessing

Sequencing reads were preprocessed using the nf-core/rnaseq (v.1.2) pipeline[58]. Read quality was assessed using FastQC (v.0.11.8)[59]. Trim Galore (v.0.5.0)[60] was used to trim sequencing reads, eliminating the remains of Illumina adaptors and discarding reads that were shorter than 20 bp. The resulting reads were mapped onto the mouse genome (GRCm38, Ensembl[61] release 81) using HiSAT2 (v.2.1.0)[62] and quantified using featureCounts (v.1.6.2)[63]. Reads per kilobase per million mapped reads (RPKM) and transcripts per million (TPM) gene expression values were calculated from the trimmed mean of *M*-values (TMM)-normalized counts per million (CPM) values using the Bioconductor package edgeR (v.3.30.0)[64] and R (v.4.0.0)[65]. Differential gene expression analysis and PCA were performed using the

Bioconductor package DESeq2 (v.1.28.1)[66]. Variance-stabilizing transformation of count data was applied to visualize the sample-to-sample distances in PCA. Genes were considered to be differentially expressed if showed an adjusted $P < 0.05$.

Functional profiling of cell subpopulations

Functional enrichment analysis of the subsets of differentially expressed genes was performed using g:Profiler web server[67] with the g:SCS significance threshold, 'Only annotated' statistical domain scope, and canonical pathway KEGG[68], Reactome[69] and Wiki Pathways[70] sets. For each gene subset, the top five significant gene sets were selected for representation.

GSEA

The RPKM matrix after the removal of low-count genes (edgeR (v.3.30.0)[64]) was used as an input for the GSEA (v.4.0.3) software[71]. We used the signal-to-noise metric to rank the genes, 1,000 permutations with the gene set permutation type and weighted enrichment statistics. Gene set sizes were chosen as 15–500 for MSigDB 7.0 GO:BP and 10–1,000 for MSigDB 7.0 canonical pathways (BioCarta, KEGG, PID, Reactome and WikiPathways)[72]. Gene sets passing the FDR < 0.25 threshold were processed for further analysis. Network representation and clustering of GSEA results were performed using EnrichmentMap (v.3.2.1)[73] and AutoAnnotate (v.1.3.2)[74] for Cytoscape (v.3.7.2)[75] with the Jaccard coefficient set to 0.25.

Functional profiling of SASP

We checked whether upregulated genes (DESeq2 adjusted $P < 0.05$ and $\log_2$[fold change] > 0) from each Sen versus NSen comparison can be expressed in a form of secreted proteins by combining the evidence from multiple data sources: GO[76] cellular component (GO:CC), Uniprot[77], VerSeDa[78], Human Protein Atlas[79] and experimental data reporting

SASP[10,80]. The genes encoding extracellular (GO:CC) and/or secreted (other sources) products, with evidence from at least one source, were included in the final list of SASP genes (1,912 in total). Functional enrichment analysis was performed using the g:Profiler web server[67] with the g:SCS significance threshold, ‘Only annotated’ statistical domain scope, and canonical pathway sets from KEGG, Reactome and Wiki Pathways. Gene sets passing the FDR < 0.05 threshold were processed for further analysis. Network representation and clustering of the g:Profiler results were performed using EnrichmentMap (v.3.2.1) and AutoAnnotate (v.1.3.2) for Cytoscape (v.3.7.2) with the Jaccard coefficient set to 0.25.

Comparative enrichment analysis of senescent cells and previously published ageing datasets

We used the minimum hypergeometric test implemented in the R package mHG (v.1.1)[81] for the comparative enrichment analysis of senescent cells and previously published ageing datasets: mouse[36], rat (Gene Expression Omnibus (GEO): GSE53960), African turquoise killifish (GEO: GSE69122), and human (GTEx[82] v6p). Data processing and analysis were performed as described previously[36].

scRNA-seq and analysis

scRNA-seq was performed using the Chromium Single Cell 3′ GEM, Library & Gel Bead Kit v3, 16 rxns (10x Genomics, PN-1000075) according to the manufacturer’s instructions and targeting a recovery of 5,000 cells per dataset. Each dataset was obtained with a sample size of two mouse biological replicates. The libraries were constructed as instructed in the manufacturer’s protocol and sequenced using the MGI DNBSEQ-T7 sequencer platform. The average read depth across the samples was 15,551 per cell. Sequencing reads were processed with STARsolo (v.2.7.3a)[83] using the mouse reference genome mm10 (GENCODE vM23 (ref. [84])).

From the filtered barcode and count matricesm, downstream analysis was performed using R (v.4.0.3). Quality control, filtering, data clustering,

visualization and differential expression analysis were performed using the Seurat (v.4.0.3) and DoubletFinder (v.2.0) R packages[85,86]. Datasets were processed following Seurat standard integration protocol according to the tutorial instructions. Genes expressed in less than 3 cells and cells with fewer than 500 features, less than 2,000 transcripts and more than 20% reads mapping to mitochondrial genes as well as cells identified as doublets by DoubletFinder were removed. PCA was performed for dimensionality reduction and the first 30 components were used for UMAP embedding and clustering.

ATAC-seq sample and library preparation

Omni-ATAC-seq was performed in freshly sorted cells as described previously[87,88]. After the transposition reaction and purification, the transposed fragments were amplified using 50 µl of PCR mix (20 µl of DNA, 2.5 µl of custom Nextera PCR primers 1 and 2, and 25 µl of KAPA HiFi HS Ready Mix for a total of 15 cycles). The PCR amplification conditions were as follows: 72 °C for 5 min; 95 °C for 30 s; 15 cycles of 95 °C for 10 s, 63 °C for 30 s and 72 °C for 60 s; and a final extension at 72 °C for 5 min. After PCR amplification, the libraries were purified, and the size was selected from 150 to 800 bp using AMPure XP beads. Paired-end sequencing was performed with 50 cycles on the Illumina NovaSeq 6000 platform.

Bulk ATAC-seq data preprocessing

Read quality was assessed using FastQC (v.0.11.8). All adaptors were removed using Fastp (v.0.21.0)[89]. The clean reads were then aligned to mm10 mouse genome assembly using Bowtie2 (v.2.2.5)[90] with the settings '--very sensitive'. Low-mapping-quality reads were removed using samtools (v.1.3.1)[91] with the settings '-q 30'. BigWig files were generated using deeptools (v.3.3.1)[92] with the settings '-normalizeUsing CPM'. Peaks were called using Macs2 (v.2.1.0)[93] with the options '--nomodel --keep-dup -q 0.01'. For differential accessibility analysis, union peak sets were created using Bedtools (v.2.29.2)[94], reads corresponding to each region were assigned by FeatureCounts. Differentially accessible peaks were identified

using DESeq2 (v.1.24.0) with the criteria of adjusted $P < 0.1$ and an absolute value of $\log_2$[fold change] > 1. Differentially accessible peaks were further annotated by HOMER (v.4.10.4)[95], the associated motif enrichment analysis was performed by HOMER using the default settings.

Analysis of senescence-induced changes in promoter chromatin accessibility

An *MA* plot ($\log_2$-transformed fold change versus mean average) was used to visualize changes in chromatin accessibility for all peaks. As a peak score, we used an average of TPM-normalized read counts: (1) reads per kilobase were calculated by division of the read counts by the length of each peak in kilobases; (2) the per million scaling factor was calculated as a sum of all reads per kilobase for each sample; (3) reads per kilobase were divided by the per million scaling factor; (4) peaks with the 'promoter-TSS' annotation TSS ± 1kb were selected and the average was calculated for each group. For *MA* plots, we included only those peaks with an average normalized signal > 5. The number of peaks with a $\log_2$-transformed fold change of >1 or <−1 was calculated. Normalized ATAC-seq signal profiles of proximal promoters were visualized for key genes using the Integrative Genomic Viewer (v.2.8.13)[96].

Transcription factor analysis and activity prediction

For the analysis of transcription regulation, we combined the results of several methods: (1) motif enrichment analysis of differentially expressed genes with the TRANSFAC_and_JASPAR_PWMs and ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X libraries using the R package EnrichR (v.2.1)[97]; (2) upstream regulator analysis of differentially expressed genes using the commercial Ingenuity Pathway Analysis (IPA, QIAGEN) software[98]; (3) analysis of transcription factor differential expression using DESeq2 (v.1.28.1); (4) motif enrichment analysis of differentially accessible regions using HOMER (v.4.10.4).

Potential regulators from EnrichR and IPA results passing the threshold of *P*

< 0.05 were used to build a union set of transcription factors, which was further filtered to retain only the molecules with DESeq2 baseMean value > 0. For further validation of the activity status, transcription factors were matched to the known HOMER motifs passing the Benjamini $Q < 0.05$ threshold.

A discrete scoring scale (inhibited, possibly inhibited, unknown/contradictory, possibly activated, activated) was used to evaluate transcription factor activity based on combined evidence from the EnrichR, IPA, DESeq2 and HOMER results. We used *z*-score statistics to define the activity status of transcription factors from the EnrichR analysis results by matching the differential expression of target genes with activatory and inhibitory interactions from the Bioconductor package DoRothEA (v.1.0.0)[99] and the web-based TRRUST v.2 database[100]. To define the activity status of transcription factors from IPA upstream regulators analysis results, IPA-calculated *z*-score and analysis bias was taken into account. Activity predictions were further corrected by differential expression of transcription factors using DESeq2. The expression *z*-score statistical value was calculated to functionally classify transcription factors as activators or repressors on the basis of the proportion of upregulated and downregulated target genes. We further calculated the chromatin accessibility *z*-score to estimate the prevalence of HOMER motif enrichment in open versus closed regions that together with the predicted transcription factor function enabled us to validate the RNA-seq activity predictions using ATAC-seq data.

To estimate the level of confidence, for each enrichment result, we calculated a discrete 'trust' score, with each point assigned for: (1) EnrichR adjusted $P < 0.05$; (2) IPA $P < 0.05$; (3) activity status 'activated' or 'inhibited'; (4) unidirectional absolute *z*-scores of >2 from both the EnrichR and IPA results; (5) concordance between transcription factor differential expression and the prediction of its activity score; (6) activity validated by the analysis of ATAC-seq data. Transcription factors with average trust > 1 were processed for further analysis.

Functional profiling of transcription factor target gene regulation

For each transcription factor, we merged the target genes from EnrichR and IPA results, split them into upregulated and downregulated and processed them to functional enrichment analysis of canonical pathways (KEGG, Reactome) and GO:BP using R package gprofiler2 (v.0.1.9)[101] with the following parameters: correction method 'FDR', 'custom_annotated' domain score consisting of target genes for all studied transcription factors. Electronic GO annotations were excluded. Gene sets that passed the FDR < 0.05 threshold were processed for further analysis. For GO:BP, we selected the gene sets with a term size of >15 and <500 genes. Transcription factors were further mapped based on the matching terms from the gprofiler2 results to the main functional clusters of the gene sets created previously in GSEA/Cytoscape analysis.

Transcription factors mapped to the same functional cluster in ≥8 (out of 12) Sen versus NSen comparisons were processed for further filtering. We scored as 1 point in each case when any of the following attributes had a value above the upper quartile for a given cluster: number of comparisons, percentage of GSEA terms among all terms with enrichment, $-\log_{10}$ of the average minimum FDR and average trust score. Moreover, we scored as 1 point if the transcription factor was associated with senescence in literature. For graphical representation, we selected examples of transcription factors and target genes based on literature research: 19 transcription factors (out of 29 with a score of ≥2) mapped to 9 clusters (matrix remodelling/fibrosis, interferon signalling, chemotaxis, lipid uptake, IGF regulation, detoxification, gene expression and protein translation, cell cycle, and DNA repair).

Functional profiling of transcriptional regulation of SASP

For each transcription factor upregulated, target genes from the EnrichR and IPA results were merged and intersected with the list of SASP genes. For SASP genes, we extracted GO:MF terms, clustered them into 12 categories (adhesion molecule, chemokine, complement component, cytokine, enzyme, enzyme regulator, extracellular matrix constituent, growth factor, hormone, ligand, proteinase and receptor) and estimated the enrichment of GO:MF clusters with a hypergeometric test using the R function phyper. Correction

for multiple comparisons was performed using the Benjamini–Hochberg procedure.

Transcription factors that had target enrichment in the same GO:MF cluster with $P < 0.05$ in ≥8 (out of 12) Sen versus NSen comparisons were processed for further filtering. We scored 1 point in each case in which any of the following attributes had a value above the upper quartile for a given cluster: the percentage of secreted proteins among targets, the number of comparisons with $P < 0.05$, the number of comparisons with adjusted $P < 0.05$, $-\log_{10}$ of the average P value, the average trust score. Moreover, we used ATAC-seq data analysis to score 1 point in cases in which the transcription factor motif was present in the promoter region of at least one SASP gene within the cluster. For graphical representation, we selected 17 transcription factors with a score of ≥2 and with ≥3 comparisons with adjusted $P < 0.05$. They were associated with five GO:MF categories (extracellular matrix constituent, cytokine, chemokine, complement component and growth factor), for which we selected the most common target genes.

Analysis of lipid metabolism gene set

For the analysis of lipid metabolism, we constructed a gene set using data from multiple sources: KEGG pathway maps (fatty acid degradation, cholesterol metabolism, regulation of lipolysis in adipocytes), WikiPathways (fatty acid oxidation, fatty acid beta oxidation, mitochondrial LC-fatty acid beta-oxidation, fatty acid omega oxidation, fatty acid biosynthesis, triacylglyceride synthesis, sphingolipid metabolism (general overview), sphingolipid metabolism (integrated pathway), cholesterol metabolism (includes both Bloch and Kandutsch–Russell pathways) and cholesterol biosynthesis) and literature research[102,103,104]. We further estimated the expression of these genes by filtering DESeq2 results (adjusted $P < 0.05$ in at least 3 out of 12 comparisons) and extracted $\log_2$-transformed fold change values to plot the difference in expression between senescent and non-senescent cells.

Reconstruction of ligand–receptor mediated cell–cell communication networks

For reconstructing cell–cell communication networks, we modified the single-cell-based method, FunRes, to account for bulk gene expression profiles[39]. In brief, transcription factors with an expression value of more than 1 TPM were considered to be expressed. Receptors regulating these transcription factors were detected using a Markov chain model of signal transduction to detect high-probability intermediate signalling molecules[105]. Ligand–receptor interactions between two cell populations were reconstructed if (1) the receptor is expressed and regulates any transcription factor, (2) the ligand is expressed and (3) the receptor–ligand interaction is contained in the cell–cell interaction scaffold. Finally, a score is assigned to every interaction by multiplying the average receptor and ligand expression in their respective cell populations. Significance was assessed by permuting cell population labels 100 times and recomputing the interaction scores in the permuted datasets. Interactions were considered to be significant if they were at least 2 s.d. greater than the mean of the permuted interaction scores. Only significant interactions were retained in the final network.

Downstream analysis of senescence-induced ligand–receptor interactions

For the functional profiling, we selected ligand–receptor interactions between three senescent cell populations (SCs, FAPs and MCs) and a non-senescent SC population in old mice at 3 d.p.i. We used the Bioconductor package SPIA (v.2.40.0)[40] with a reduced set of non-disease KEGG pathway maps to evaluate the activity of a pathway's downstream ligand–receptor interactions. For each interaction, differentially expressed target transcription factors in non-senescent SCs were split into upregulated and downregulated in comparison to senescent SCs. As a reference set of genes, we took a list of target transcription factors from all of the interactions studied. SPIA analysis was performed with 2,000 permutations, and pPERT and pNDE were combined using the Fisher's product method. Pathways passing the pGFdr <

0.05 threshold were considered to be significantly enriched. For each pathway, we calculated the ratio of ligand–receptor interactions that activate or inhibit the pathway to the total number of interactions analysed. For results representation, we selected eight activated and eight inhibited pathways with the highest ratio of interactions.

Statistical analysis

The sample size of each experimental group is described in the corresponding figure caption, and all of the experiments were conducted with at least three biological replicates unless otherwise indicated. GraphPad Prism was used for all statistical analyses except for sequencing-data analysis. Quantitative data displayed as histograms are expressed as mean ± s.e.m. (represented as error bars). Results from each group were averaged and used to calculate descriptive statistics. Mann–Whitney *U*-tests (independent samples, two-tailed) were used for comparisons between groups unless otherwise indicated. $P < 0.05$ was considered to be statistically significant. Experiments were not randomized.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The bulk RNA-seq, scRNA-seq and ATAC-seq data supporting the findings of this study have been deposited at the GEO under accession number GSE196613. Source data are provided with this paper.

References

1. Fuchs, E. & Blau, H. M. Tissue stem cells: architects of their niches. *Cell Stem Cell* **27**, 532–556 (2020).

2. Sousa-Victor, P., Garcia-Prat, L. & Munoz-Canoves, P. Control of

satellite cell function in muscle regeneration and its disruption in ageing. *Nat. Rev. Mol. Cell Biol.* **23**, 204–226 (2022).

3. Roy, A. L. et al. A blueprint for characterizing senescence. *Cell* **183**, 1143–1146 (2020).

4. Franceschi, C. & Campisi, J. Chronic inflammation (inflammaging) and its potential contribution to age-associated diseases. *J. Gerontol. A* **69**, S4–S9 (2014).

5. Sousa-Victor, P. et al. Geriatric muscle stem cells switch reversible quiescence into senescence. *Nature* **506**, 316–321 (2014).

6. Garcia-Prat, L. et al. Autophagy maintains stemness by preventing senescence. *Nature* **529**, 37–42 (2016).

7. Ho, T. T. et al. Autophagy maintains the metabolism and function of young and old stem cells. *Nature* **543**, 205–210 (2017).

8. Brunet, A., Goodell, M. A. & Rando, T. A. Ageing and rejuvenation of tissue stem cells and their niches. *Nat. Rev. Mol. Cell Biol.* https://doi.org/10.1038/s41580-022-00510-w (2022).

9. Hong, X. et al. Mitochondrial dynamics maintain muscle stem cell regenerative competence throughout adult life by regulating metabolism and mitophagy. *Cell Stem Cell* **29**, 1298–1314 (2022).

10. Basisty, N. et al. A proteomic atlas of senescence-associated secretomes for aging biomarker development. *PLoS Biol.* **18**, e3000599 (2020).

11. Baker, D. J. et al. Naturally occurring p16^{Ink4a}-positive cells shorten healthy lifespan. *Nature* **530**, 184–189 (2016).

12. Bussian, T. J. et al. Clearance of senescent glial cells prevents tau-dependent pathology and cognitive decline. *Nature* **562**, 578–582 (2018).

13. Chang, J. et al. Clearance of senescent cells by ABT263 rejuvenates

aged hematopoietic stem cells in mice. *Nat. Med.* **22**, 78–83 (2016).

14. Childs, B. G. et al. Senescent intimal foam cells are deleterious at all stages of atherosclerosis. *Science* **354**, 472–477 (2016).

15. Xu, M. et al. Senolytics improve physical function and increase lifespan in old age. *Nat. Med.* **24**, 1246–1256 (2018).

16. Ng, P. Y., McNeely, T. L. & Baker, D. J. Untangling senescent and damage-associated microglia in the aging and diseased brain. *FEBS J.* https://doi.org/10.1111/febs.16315 (2021).

17. Zhu, Y. et al. The Achilles' heel of senescent cells: from transcriptome to senolytic drugs. *Aging Cell* **14**, 644–658 (2015).

18. Demaria, M. et al. An essential role for senescent cells in optimal wound healing through secretion of PDGF-AA. *Dev. Cell* **31**, 722–733 (2014).

19. Krizhanovsky, V. et al. Senescence of activated stellate cells limits liver fibrosis. *Cell* **134**, 657–667 (2008).

20. Munoz-Espin, D. et al. Programmed cell senescence during mammalian embryonic development. *Cell* **155**, 1104–1118 (2013).

21. Storer, M. et al. Senescence is a developmental mechanism that contributes to embryonic growth and patterning. *Cell* **155**, 1119–1130 (2013).

22. Chiche, A. et al. Injury-induced senescence enables in vivo reprogramming in skeletal muscle. *Cell Stem Cell* **20**, 407–414 (2017).

23. Mosteiro, L. et al. Tissue damage and senescence provide critical signals for cellular reprogramming in vivo. *Science* **354**, aaf4445 (2016).

24. Ocampo, A. et al. In vivo amelioration of age-associated hallmarks by partial reprogramming. *Cell* **167**, 1719–1733 (2016).

25. Zhang, J. et al. A novel near-infrared fluorescent probe for sensitive detection of β-galactosidase in living cells. *Anal. Chim. Acta* **968**, 97–

104 (2017).

26. Debacq-Chainiaux, F., Erusalimsky, J. D., Campisi, J. & Toussaint, O. Protocols to detect senescence-associated beta-galactosidase (SA-βgal) activity, a biomarker of senescent cells in culture and in vivo. *Nat. Protoc.* **4**, 1798–1806 (2009).

27. Serrano, A. L. & Munoz-Canoves, P. Fibrosis development in early-onset muscular dystrophies: mechanisms and translational implications. *Semin. Cell Dev. Biol.* **64**, 181–190 (2017).

28. Prieto, L. I., Graves, S. I. & Baker, D. J. Insights from in vivo studies of cellular senescence. *Cells* **9**, 954 (2020).

29. Moiseeva, V. et al. Context-dependent roles of cellular senescence in normal, aged, and disease states. *FEBS J.* https://doi.org/10.1111/febs.16573 (2022).

30. Wree, A. et al. NLRP3 inflammasome activation is required for fibrosis development in NAFLD. *J. Mol. Med.* **92**, 1069–1082 (2014).

31. Chan, M. et al. Novel insights from a multiomics dissection of the Hayflick limit. *eLife* **11**, e70283 (2022).

32. Tilstra, J. S. et al. NF-κB inhibition delays DNA damage-induced senescence and aging in mice. *J. Clin. Invest.* **122**, 2601–2612 (2012).

33. Kuilman, T. et al. Oncogene-induced senescence relayed by an interleukin-dependent inflammatory network. *Cell* **133**, 1019–1031 (2008).

34. Kandhaya-Pillai, R. et al. TNFalpha-senescence initiates a STAT-dependent positive feedback loop, leading to a sustained interferon signature, DNA damage, and cytokine secretion. *Aging* **9**, 2411–2435 (2017).

35. Meng, X. M., Nikolic-Paterson, D. J. & Lan, H. Y. TGF-β: the master regulator of fibrosis. *Nat. Rev. Nephrol.* **12**, 325–338 (2016).

36. Benayoun, B. A. et al. Remodeling of epigenome and transcriptome landscapes with aging in mice reveals widespread induction of inflammatory responses. *Genome Res.* **29**, 697–709 (2019).

37. Stegeman, R. & Weake, V. M. Transcriptional signatures of aging. *J. Mol. Biol.* **429**, 2427–2437 (2017).

38. Shavlakadze, T. et al. Age-related gene expression signature in rats demonstrate early, late, and linear transcriptional changes from multiple tissues. *Cell Rep.* **28**, 3263–3273 (2019).

39. Jung, S., Singh, K. & Del Sol, A. FunRes: resolving tissue-specific functional cell states based on a cell-cell communication network model. *Brief. Bioinform.* **22**, bbaa283 (2020).

40. Tarca, A. L. et al. A novel signaling pathway impact analysis. *Bioinformatics* **25**, 75–82 (2009).

41. Nazir, S. et al. Interaction between high-density lipoproteins and inflammation: function matters more than concentration! *Adv. Drug Deliv. Rev.* **159**, 94–119 (2020).

42. Chong, M. et al. CD36 initiates the secretory phenotype during the establishment of cellular senescence. *EMBO Rep.* **19**, e45274 (2018).

43. Saitou, M. et al. An evolutionary transcriptomics approach links CD36 to membrane remodeling in replicative senescence. *Mol. Omics* **14**, 237–246 (2018).

44. Yousefzadeh, M. J. et al. An aged immune system drives senescence and ageing of solid organs. *Nature* **594**, 100–105 (2021).

45. Hernandez-Segura, A. et al. Unmasking transcriptional heterogeneity in senescent cells. *Curr. Biol.* **27**, 2652–2660 (2017).

46. Kang, C. et al. The DNA damage response induces inflammation and senescence by inhibiting autophagy of GATA4. *Science* **349**, aaa5612 (2015).

47. Vizioli, M. G. et al. Mitochondria-to-nucleus retrograde signaling drives formation of cytoplasmic chromatin and inflammation in senescence. *Genes Dev.* **34**, 428–445 (2020).

48. West, A. P. et al. Mitochondrial DNA stress primes the antiviral innate immune response. *Nature* **520**, 553–557 (2015).

49. De Cecco, M. et al. L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**, 73–78 (2019).

50. Saito, Y., Chikenji, T. S., Matsumura, T., Nakano, M. & Fujimiya, M. Exercise enhances skeletal muscle regeneration by promoting senescence in fibro-adipogenic progenitors. *Nat. Commun.* **11**, 889 (2020).

51. Suelves, M. et al. uPA deficiency exacerbates muscular dystrophy in MDX mice. *J. Cell Biol.* **178**, 1039–1051 (2007).

52. Grounds, M. D., Sorokin, L. & White, J. Strength at the extracellular matrix-muscle interface. *Scand. J. Med. Sci. Sports* **15**, 381–391 (2005).

53. Segales, J. et al. Sestrin prevents atrophy of disused and aging muscles by integrating anabolic and catabolic signals. *Nat. Commun.* **11**, 189 (2020).

54. Sacco, A. et al. Short telomeres and stem cell exhaustion model Duchenne muscular dystrophy in mdx/mTR mice. *Cell* **143**, 1059–1071 (2010).

55. Le, G., Lowe, D. A. & Kyba, M. Freeze injury of the tibialis anterior muscle. *Methods Mol. Biol.* **1460**, 33–41 (2016).

56. Fumagalli, M. et al. Telomeric DNA damage is irreparable and causes persistent DNA-damage-response activation. *Nat. Cell Biol.* **14**, 355–365 (2012).

57. Garcia-Prat, L. et al. FoxO maintains a genuine muscle stem-cell quiescent state until geriatric age. *Nat. Cell Biol.* **22**, 1307–1318 (2020).

58. Ewels, P. A. et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).

59. Andrews, S. FastQC: a quality control tool for high throughput sequence data (2010); http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

60. Krueger, F. Trim Galore (2021); https://github.com/FelixKrueger/TrimGalore.

61. Cunningham, F. et al. Ensembl 2015. *Nucleic Acids Res.* **43**, D662–D669 (2015).

62. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

63. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

64. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

65. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022); http://www.R-project.org/

66. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

67. Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).

68. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

69. Jassal, B. et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).

70. Martens, M. et al. WikiPathways: connecting communities. *Nucleic Acids Res.* **49**, D613–D621 (2021).

71. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

72. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

73. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* **5**, e13984 (2010).

74. Kucera, M., Isserlin, R., Arkhangorodsky, A. & Bader, G. D. AutoAnnotate: a Cytoscape app for summarizing networks with semantic annotations. *F1000Research* **5**, 1717 (2016).

75. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

76. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

77. UniProt, C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).

78. Cortazar, A. R., Oguiza, J. A., Aransay, A. M. & Lavin, J. L. VerSeDa: vertebrate secretome database. *Database* **2017**, baw171 (2017).

79. Thul, P. J. et al. A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).

80. Freund, A., Orjalo, A. V., Desprez, P. Y. & Campisi, J. Inflammatory networks during cellular senescence: causes and consequences. *Trends*

Mol. Med. **16**, 238–246 (2010).

81. Perl, K. mHG: minimum-hypergeometric test. R package version 1.0 (2015); https://cran.r-project.org/web/packages/mHG/index.html

82. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).

83. Kaminow, B., Yunusov, D. & Dobin, A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. Preprint at *bioRxiv* https://doi.org/10.1101/2021.05.05.442755 (2021).

84. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).

85. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).

86. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337 (2019).

87. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).

88. Picelli, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).

89. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

90. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

91. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

92. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).

93. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome biology **9**, R137 (2008).

94. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

95. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

96. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

97. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).

98. Kramer, A., Green, J., Pollard, J. Jr & Tugendreich, S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* **30**, 523–530 (2014).

99. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **29**, 1363–1375 (2019).

100. Han, H. et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* **46**, D380–D386 (2018).

101. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2—an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research* **9**, ELIXIR-

709 (2020).

102. Flor, A. C., Wolfgeher, D., Wu, D. & Kron, S. J. A signature of enhanced lipid metabolism, lipid peroxidation and aldehyde stress in therapy-induced senescence. *Cell Death Discov.* **3**, 17075 (2017).

103. Stahl, A. A current review of fatty acid transport proteins (SLC27). *Pflugers Arch.* **447**, 722–727 (2004).

104. Zani, I. A. et al. Scavenger receptor structure and function in health and disease. *Cells* **4**, 178–201 (2015).

105. Ravichandran, S., Hartmann, A. & Del Sol, A. SigHotSpotter: scRNA-seq-based computational tool to control cell subpopulation phenotypes for cellular rejuvenation strategies. *Bioinformatics* **36**, 1963–1965 (2019).

Acknowledgements

We thank M. Jardí, A. Navarro, J. M. Ballestero, K. Slobodnyuk, M. González, J. López and M. Raya for their technical contributions; A. Harada and K. Tanaka for assistance in ATAC-seq; all of the members of the P.M.-C. laboratory for discussions; J. Campisi for p16-3MR mice; J. A. Fernández-Blanco (PRBB Animal Facility); O. Fornas (UPF/CRG FACS Facility); E. Rebollo (IBMB Molecular Imaging Platform); V. A. Raker for manuscript editing; and the members of the Myoage network (A. Maier) for human material. We acknowledge funding from MINECO-Spain (RTI2018-096068, to P.M.-C. and E.P.); ERC-2016-AdG-741966, LaCaixa-HEALTH-HR17-00040, MDA, UPGRADE-H2020-825825, AFM, DPP-Spain, Fundació La MaratóTV3-80/19-202021 and MWRF to P.M.-C.; Fundació La MaratóTV3-137/38-202033 to A.L.S.; María-de-Maeztu Program for Units of Excellence to UPF (MDM-2014-0370) and Severo-Ochoa Program for Centers of Excellence to CNIC (SEV-2015-0505). This work was also supported by JST-CREST JPMJCR16G1 and MEXT/JSPS JP20H00456/18H05527 to Y.O.; the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16030502) to M.A.E.; V.M. and A.C. were supported by FPI and Maria-de-Maeztu predoctoral fellowships, respectively, and V.S. by a

Marie Skłodowska-Curie individual fellowship. Parts of the figures were drawn using pictures from Servier Medical Art. Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licences/by/3.0/).

Author information

Authors and Affiliations

1. Department of Medicine and Life Sciences, Pompeu Fabra University, Barcelona, Spain

   Victoria Moiseeva, Andrés Cisneros, Valentina Sica, Oleg Deryagin, Eva Andrés, Jessica Segalés, Laura Ortet, Vera Lukesova, Antonio L. Serrano, Eusebio Perdiguero & Pura Muñoz-Cánoves

2. CIBERNED, Barcelona, Spain

   Victoria Moiseeva, Andrés Cisneros, Valentina Sica, Oleg Deryagin, Eva Andrés, Jessica Segalés, Laura Ortet, Vera Lukesova, Antonio L. Serrano, Eusebio Perdiguero & Pura Muñoz-Cánoves

3. Laboratory of Integrative Biology, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China

   Yiwei Lai, Juan An, Giacomo Volpe & Miguel A. Esteban

4. Key Laboratory of Regenerative Biology, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China

   Yiwei Lai, Juan An, Giacomo Volpe & Miguel A. Esteban

5. Guangdong Provincial Key Laboratory of Stem Cells and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese

Academy of Sciences, Guangzhou, China

Yiwei Lai, Juan An, Giacomo Volpe & Miguel A. Esteban

6. CIC bioGUNE-BRTA (Basque Research and Technology Alliance), Bizkaia Technology Park, Derio, Spain

Sascha Jung & Antonio del Sol

7. University of Science and Technology of China, Hefei, China

Juan An

8. Genomic Unit, Centro Nacional de Investigaciones Cardiovasculares and CIBERCV, Madrid, Spain

Alberto Benguria & Ana Dopazo

9. ICREA, Barcelona, Spain

Salvador Aznar-Benitah & Pura Muñoz-Cánoves

10. Institute for Research in Biomedicine and BIST, Barcelona, Spain

Salvador Aznar-Benitah

11. Laboratory of Chemistry & Biology, Graduate School of Pharmaceutical Sciences and School of Medicine, The University of Tokyo, Tokyo, Japan

Yasuteru Urano

12. Computational Biology Group, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Antonio del Sol

13. IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Antonio del Sol

14. Bioland Laboratory, Guangzhou Regenerative Medicine and Health Guangdong Laboratory, Guangzhou, China

Miguel A. Esteban

15. Division of Transcriptomics. Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan

Yasuyuki Ohkawa

16. Altos labs Inc, San Diego, CA, USA

Antonio L. Serrano, Eusebio Perdiguero & Pura Muñoz-Cánoves

17. Cardiovascular Regeneration Program, CNIC Centro Nacional de Investigaciones Cardiovasculares, Madrid, Spain

Pura Muñoz-Cánoves

Contributions

Conceptualization: P.M.-C., E.P. and V.M. Formal analysis: V.M., A.C., O.D., S.J., A.L.S. and E.P. Investigation: V.M., A.C., V.S., L.Y., S.J., E.A., J.A., J.S., L.O., V.L., G.V., A.B., A.D., S.A.-B., Y.U., A.d.S., M.A.E., Y.O., A.L.S. and E.P. Writing—original draft: P.M.-C., V.M. and E.P. Writing—review and editing: P.M.-C., V.M., A.C., A.L.S. and E.P. Funding acquisition: P.M.-C. Resources: Y.U. Visualization: V.M., A.C., A.L.S., E.P. and P.M.-C. Supervision: P.M.-C. and E.P.

Corresponding authors

Correspondence to Eusebio Perdiguero or Pura Muñoz-Cánoves.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Tamara Tchkonia and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Characteristics of senescent cells in regenerating muscle of young and old mice.

a) Quantification of Renilla luciferase activity in regenerating muscle from young and old mice at the indicated d.p.i. The luciferase activities are normalized to the activity of basal young muscle (*n* = 6 mice for young basal, 8 muscles from 6 mice for young 3 d.p.i., 8 muscles from 4 mice for young 7 d.p.i., 7 muscles from 5 mice for young 10 d.p.i., 5 mice for old basal, 5 muscles from 4 mice for old 3 d.p.i., 6 muscles from 4 mice for old 7 d.p.i., and 7 muscles from 4 mice for old 10 d.p.i.). **b**) Representative images and quantification of CSA and frequency distribution analysis of $MYH3^+$ fibre size in TA muscles from young and old mice at 7 d.p.i. (*n* = 9 TA from 6 mice for young and 5 mice for old). **c**) RT-qPCR of *mRFP*, *Cdkn2a*, and $p19^{ARF}$ in young and old muscle tissue from p16-3MR mice at the indicated d.p.i. (*n* = 4 mice in *mRFP* and *Cdkn2a* for old 3 d.p.i., 5 TA from 3 mice for old 7 d.p.i., 5 TA from 4 mice for young 7 d.p.i., 5 mice in *mRFP*, for

young basal and 3 d.p.i. and *Cdkn2a* for young 3 d.p.i., 6 TA from 5 mice in *p19^{ARF}* for old 3 d.p.i., 6 mice in *Cdkn2a* and *p19^{ARF}* for young basal and *p19^{ARF}* for young 3 d.p.i., and 8 mice for old basal). **d**) Freshly isolated SCs were obtained from skeletal muscle tissue and cultured in the presence of etoposide (1 μM) or DMSO for 4 days. Cells were stained with SPiDER, C12FDG, or SA-β-gal staining in parallel and analysed by flow cytometry or microscopy to assess their entry into senescence ($n = 4$ mice/group). Unstained samples were used to determine the threshold for C12FDG and SPiDER populations. Histogram representation of SPiDER intensity, representative images of SA-β-gal and quantification are shown. **e**) Histogram representation of SPiDER-β-gal staining and gating strategy employed for isolation of SPiDER^{Low}, SPiDER^{Medium}, and SPiDER^{High} populations from injured skeletal muscle at 3 d.p.i. **f**) Representative images and quantification of SA-β-gal staining ($n = 4$ mice/group), cell area ($n = 205$ SPiDER^{Low} cells, 168 SPiDER^{Medium} cells, and 220 SPiDER^{High} cells), and lamin B1 expression ($n = 36$ SPiDER^{Low} cells, 33 SPiDER^{Medium} and SPiDER^{High} cells) in freshly sorted SPiDER^{Low}, SPiDER^{Medium}, and SPiDER^{High} from regenerating muscles at 3 d.p.i. **g**) Graphs representing multimarker senescence scoring in freshly sorted SPiDER^{+} population from 3 d.p.i. regenerating tissue of young mice. Scoring was calculated with cell size, SA-β-gal, lamin B1, and proliferation rate (Fig. 1e and Extended Data Fig. 1f; see Methods). **h**) Gating strategy used to isolate SPiDER^{+} cells from regenerating muscle tissue of young mice at 3 d.p.i. Cells were divided into two major populations with anti-CD45 antibodies to overcome differences in auto-fluorescence of hematopoietic and non-hematopoietic populations. Fluorescence Minus One (FMO) and samples from non-injured muscle tissue were used to set the threshold for SPiDER^{+} staining within each cell population. **i**) Representative immunofluorescence images showing CDKN2A^{+} and γH2Ax^{+} cells in proximity to SCs, FAPs and MCs identified with anti-Pax7, anti-PDGFRα and anti-F4/80 antibodies, respectively, in regenerating muscles of young mice at 4 d.p.i.. Arrows indicate CDKN2A^{+} and γH2Ax^{+} cells. **j**) Representative images and quantification of CDKN2A^{+} cells ($n = 4$ mice/group) and γH2Ax^{+} cells ($n = 5$ mice for SCs, 3 mice for FAPs, and 4 mice for MCs) in regenerating muscles from young mice at 4 d.p.i. Each cell type was labelled with indicated antibodies and nuclei with 4,6-diaminido-2-phenylindole (DAPI). Arrows indicate CDKN2A^{+} cells. **k**)

Representative images and quantification of 53BP1$^+$ cells in regenerating human muscle. Each cell type was labelled with indicated antibodies, and nuclei with DAPI ($n = 4$ samples/group, from persons aged 81±7.5 years old). Scale bars: 50 µm (**b**); 10 µm (**d**, **f**, **i**, **j**, and **k** low magnification) and 5 µm (**k** high magnification). Results are displayed as means±s.e.m; *P* values were calculated by Mann–Whitney *U*-test (**a** between ages, **c** between ages, **b** and **d**), and Tukey's test (**a** between d.p.i., **c** between d.p.i., **f**, **j**, and **k**).

Source data

Extended Data Fig. 2 Removal of senescent cells improves skeletal muscle regeneration after acute damage throughout life.

Young and old p16-3MR mice were subjected to CTX injury, treated with vehicle or GCV during the course of regeneration and analysed at 7 d.p.i. **a**) Representative images of Sirius Red staining and quantification of SA-β-gal$^+$ cells in damaged area ($n = 5$ TA from 4 mice for vehicle and 6 TA from 3 mice for GCV), Sirius Red staining ($n = 4$ TA from 3 mice for vehicle and 5 TA from 3 mice for GCV), and frequency distribution analysis of positive MYH3 fibre size ($n = 3$ mice/group) in cryosections from old p16-3MR mice. **b**) Quantification of *in vivo* Renilla luminescence activity ($n = 6$ TA from 3 mice for vehicle and 10 TA from 5 mice for GCV), SA-β-gal$^+$ cells ($n = 6$ TA from 4 mice for vehicle and 8 TA from 4 mice for GCV), Sirius Red staining (n = 8 TA from 5 mice for vehicle and 10 mice from 5 mice for GCV), and frequency distribution analysis of positive MYH3 fibre size ($n = 4$ mice/group) in cryosections from vehicle- and GCV-treated young p16-3MR mice. **c**) Force measurements in EDL muscles of vehicle- or GCV-treated old p16-3MR mice at 10 d.p.i. ($n = 5$ EDL muscles from 4 mice for vehicle and 7 EDL from 5 mice for GCV). **d**) Force-frequency curves measured in EDL muscles of vehicle- or GCV-treated young ($n = 11$ EDL from 7 mice in vehicle and 9 EDL from 7 mice in GCV group) and old ($n = 5$ EDL from 4 mice in vehicle and 7 EDL from 5 mice in GCV group) p16-3MR mice at 10 d.p.i. **e**) Force measurements in EDL muscles of vehicle- or GCV-treated young p16-3MR mice at 10 d.p.i. ($n = 11$ EDL muscles from 8 mice for vehicle and 9 EDL from 7 mice for GCV). **f**) RT-qPCR of *Il6*, *Il1b*, *Tnf*, *Il12*,

Ccl2, *Il18*, and *Ifng* in muscle tissue from young (top, *n* = 5 TA muscles in *Tnf*, *Il12* and *Ccl2*, and 6 TA muscles in *Il6*, *Il1b*, *Il18* and *Ifng* from 4 mice for vehicle and 4 mice for GCV) and old (bottom, *n* = 4 TA muscles from 3 mice for vehicle and 6 TA muscles in *Tnf* and 5 TA in *Il6*, *Il1b*, *Il12*, *Ccl2*, *Il18*, *Ifng* from 3 mice for GCV) p16-3MR mice. **g**) Muscles of young WT mice were injured with CTX and treated with vehicle or senolytics during regeneration and analysed at 7 d.p.i. Representative images, SA-β-gal staining (*n* = 10 TA from 5 mice for vehicle and 6 TA from 3 mice for D+Q), mean CSA (*n* = 9 TA from 6 mice for vehicle and 5 TA from 3 mice for D+Q) and frequency distribution analysis (*n* = 3 mice/group) of MYH3$^+$ fibres, and Sirius Red quantification (*n* = 6 TA from 3 mice for both groups) in TA cryosections from young WT mice. **h**) Force measurements in EDL muscles of vehicle- and D+Q-treated young mice at 10 d.p.i. (*n* = 8 EDL from 5 mice in vehicle and 7 EDL from 5 mice in D+Q group). **i**) As in **g**, SA-β-gal staining, mean CSA and frequency distribution analysis of MYH3$^+$ fibres and Sirius Red quantification in TA cryosections from old WT mice (*n* = 5 mice/group) . **j**) As in **h**, force measurements in EDL muscles of vehicle- and D+Q-treated old mice at 10 d.p.i. (*n* = 5 EDL from 3 mice in vehicle and 7 EDL from 5 mice in D+Q group). **k**) Force-frequency curves measured in EDL muscles of vehicle- and D+Q-treated young (*n* = 8 EDL from 5 mice in vehicle and 7 EDL from 5 mice in D+Q group), and old (*n* = 5 EDL from 3 mice in vehicle and 7 EDL from 5 mice in D+Q group) mice at 10 d.p.i. **l**) SA-β-gal$^+$ cells and mean CSA in vehicle- or GCV-treated young p16-3MR mice after 4-days or 7-days treatment at 7 d.p.i. (*n* = 6 muscles from 4 mice for vehicle (3 to 7), 8 muscles from 4 mice GCV (3 to 7), 4 mice for vehicle (0 to 7), and 5 mice for GCV (0 to 7)). **m**) Quantification of SA-β-gal, Sirius Red staining (*n* = 5 mice for vehicle and D+Q (3 to 7) and 8 muscles from 4 mice for D+Q (0 to 7)), and mean CSA of MYH3$^+$ fibres (*n* = 5 mice for vehicle and D+Q (3 to 7) and 6 muscles from 3 mice for D+Q (0 to 7)) in vehicle- and D+Q- treated old mice after 4-days or 7-days treatment at 7 d.p.i. (*n* = 5–8 muscles from 4-5 mice). Treatments were administered from 3 to 7 d.p.i. in **b**, **g**, and **i**, from 0 to 7 d.p.i. in **a**, **b**, and **f** and from 3 to 10 d.p.i. in **c**, **d**, **e**, **h**, **j** and **k**. Scale bars 50 μm. Results are displayed as mean ± s.e.m.; *P* values were calculated by two-way ANOVA and Mixed-effects analysis (**d** and **k**) and Mann–Whitney *U*-test (**a**, **b**, **c**, **e**, **f**, **g**, **h**, **i**, **j**, **l** and **m**).

Source data

Extended Data Fig. 3 Removal of senescent cells improves skeletal muscle regeneration after microdamage and after chronic damage.

a) Quantification of *in vivo* Renilla luciferase activity in young p16-3MR mice at the indicated days after injury by micropunctures (*n* = 13 mice for Basal, 3 mice for 3 d.p.i., 5 mice for 7 d.p.i., and 4 mice for 12 d.p.i.). **b**) Representative images of SA-β-gal and MYH3 staining in cryosections of TA muscles from young WT mice at the indicated days after injury by micropunctures and quantification of SA-β-gal are shown (*n* = 5 mice for basal, 6 mice for 3 d.p.i. and 5 d.p.i., and 4 mice for 7 d.p.i.). **c**) Young p16-3MR mice were subjected to micropunctures in TA muscles and treated for 7 days with GCV, starting from the day of injury, sacrificed at 7 d.p.i. and TA muscles were analysed. Representative images of hematoxylin and eosin (H/E) staining and quantification of *in vivo* Renilla luminescence activity in TA (*n* = 6 TA from 4 mice for vehicle and 8 TA from 4 mice for GCV), SA-β-gal^{+} cells (*n* = 6 TA from 4 mice for both groups), CSA of centrally nucleated fibres and Sirius Red staining (*n* = 4 mice/group). **d**) Young WT mice were subjected to micropunctures injury in TA muscles and treated for 7 days with D+Q, starting from the day of injury. Quantification of SA-β-gal^{+} cells (*n* = 4 mice/group), CSA of centrally nucleated fibres in cryosections (*n* = 9 TA from 5 mice for vehicle and 10 TA from 5 mice for D+Q), Sirius Red staining (*n* = 8 TA from 4 mice for vehicle and 10 TA from 5 mice for D+Q). **e**) As in **c**, mRNA quantification of the indicated genes by RT-qPCR in TA muscles (*n* = 5 mice in *Il6*, *Il1b*, and *Tnf* and 6 mice in, *Il12*, *Ccl2*, and *Il18* for vehicle and 5 TA from 4 mice for GCV). **f**) Quantification of SA-β-gal^{+} cells in uninjured TA muscles from WT and mdx mice at indicated age (*n* = 9 mice WT^{3months} and mdx$^{5–7months}$, 3 mice WT$^{5–7months}$, and 8 mice mdx^{3months}). **g**) *In vivo* quantification of Renilla luminescence activity in basal muscles of p16-3MR and mdx/p16-3MR mice at indicated age (*n* = 6 mice p16-3MR^{3months} and p16-3MR$^{5–7months}$ and 5 muscles from 3 mice mdx^{3months} and mdx$^{5–7months}$). **h**) mRNA quantification of the indicated genes by RT-qPCR in TA muscles from young WT and mdx/p16-3MR mice of 5

months of age (*n* = 5 mice in *Ifng* and *Ccl2* and 6 mice for the rest of the genes for WT and 4 mice in *p19*^{ARF} and 5 mice for the rest of the genes for mdx/p16-3MR). **i**) Representative images of SA-β-gal and MYH3 staining in cryosections of TA muscles from young WT and mdx mice. **j**) Young mdx/p16-3MR mice received GCV twice a week for 2 months and were sacrificed at 5 months of age. Representative images are shown for Sirius Red staining as well as for quantification of SA-β-gal^{+} cells, Sirius Red staining, CSA and frequency distribution of regenerating fibres in TA muscles of vehicle- or GCV-treated mdx/p16-3MR mice (*n* = 6 mice for vehicle and 4 mice for GCV). **k**) Young mdx/p16-3MR mice were treated with GCV of vehicle for 2 months and force measurements were performed in EDL muscles at 5 months of age. Graphs represent maximum and specific force parameters and force-frequency curves (*n* = 9 EDL from 6 mice for vehicle and 5 EDL from 3 mice for GCV). **l**) Young mdx mice received D+Q twice a week for 2 months and muscle samples were collected at 5 months of age. Maximum and specific force parameters and force-frequency curves in EDL muscles are represented (*n* = 8 EDL muscles from 5 mice for vehicle and 7 EDL from 6 mice for D+Q). **m**) Quantifications of SA-β-gal^{+} cells, mean CSA and frequency distribution of regenerating fibres and Sirius Red staining in muscle cryosections (*n* = 5 mice/group) of vehicle- and D+Q-treated mdx mice after 2 months of treatment. Scale bars 50 µm. Results are displayed as mean ± s.e.m.; *P* values were calculated by Tukey's test (**a** and **b**), two-way ANOVA and Mixed-effects analysis in (**k** and **l** force-frequency curves) and Mann–Whitney *U*-test (**c**, **d**, **e**, **f**, **g**, **h**, **j**, **k**, **l** and **m**).

Source data

Extended Data Fig. 4 Isolation and characterization of different senescent cell types.

a) Gating strategy used to simultaneously isolate SCs, FAPs, and MCs from WT mice. Representative histogram plots from cytofluorimetric analysis are employed to assess SPiDER levels in the cell populations. FMO controls and non-injured samples were used to determine the threshold for SPiDER within each cell population. **b**) Representative pictures of Pax7, TCF4 and CD11b expression in sorted SCs, FAPs and MCs respectively. **c**) Heatmap of gene

expression levels of the indicated genes in basal, NSen and Sen SCs, FAPs and MCs. **d**) Single-cell expression levels for select gene markers. **e**) Representative images and quantification of SA-β-gal staining of freshly sorted SPiDER$^+$ and SPiDER$^-$ SCs, FAPs, and MCs from regenerating muscles at 3 d.p.i. (*n* = 3 mice/group). **f**) As in **e**, quantification of the cell area (*n* = 79 SCs^{NSen}, 55 SCs^{Sen}, 94 FAPs^{NSen}, 75 FAPs^{Sen}, 106 MCs^{NSen}, and 66 MCs^{Sen}). **g**) Representative images and quantification of lamin B1 expression of freshly sorted SPiDER$^+$ and SPiDER$^-$ SCs, FAPs, and MCs from regenerating muscles at 3 d.p.i., (*n* = 20 SCs^{NSen}, 15 SCs^{Sen}, 33 FAPs^{NSen}, 35 FAPs^{Sen}, 21 MCs^{NSen}, and 30 MCs^{Sen}; arbitrary units: AU). **h**) BrdU incorporation was quantified in cells obtained at 7 d.p.i. and cultured for 3 days (*n* = 4 mice in SCs^{NSen} and SCs^{Sen}, 7 mice in FAPs^{NSen}, and 6 mice in FAPs^{Sen}). **i**) Quantification of TUNEL assay in freshly sorted SPiDER$^+$ and SPiDER$^-$ SCs, FAPs and MCs. Cells treated with DNase were used as a positive control (*n* = 64 SCs^{NSen}, 63 SCs^{Sen}, 55 FAPs^{NSen}, 73 FAPs^{Sen}, 67 MCs^{NSen}, 56 MCs^{Sen}, 32 neg. control and 101 pos. control cells). **j**) RT-qPCR of *Cdkn2a* in freshly sorted SPiDER$^+$ and SPiDER$^-$ SCs, FAPs, and MCs from regenerating muscles at 3 d.p.i. (*n* = 4 mice/group). **k**) As in **e**, quantification of 8-oxoG (*n* = 47 SCs^{NSen}, 49 SCs^{Sen}, 60 FAPs^{NSen}, 74 FAPs^{Sen}, 65 MCs^{NSen}, and 65 MCs^{Sen}). **l**) Representative images and quantification of telomeric DDR in SPiDER$^+$ and SPiDER$^-$ sorted cells from regenerating muscle of young mice at 3 d.p.i. (*n* = 53 SCs^{NSen}, 64 SCs^{Sen}, 45 FAPs^{NSen}, 47 FAPs^{Sen}, 46 MCs^{NSen}, and 45 MCs^{Sen}). Scale bars, 10 μm (**e** and **g**), 5 μm (**l** low magnification), and 1 μm (**b** and **l** high magnification). Results are displayed as mean ± s.e.m.; *P* values from multiple *t*-tests (**e**), Tukey's test (**i**), and Mann–Whitney *U*-test (**f**, **g**, **h**, **j**, **k** and **l**).

Source data

Extended Data Fig. 5 RNAseq analysis of the different senescent cell types throughout life.

a) Scheme showing 36 different conditions (3 cell types x 12 conditions) assessed by RNA-seq. **b**) Principal component analysis (PCA) of the full transcriptome of senescent (Sen), non-senescent (NSen), and Basal SCs,

FAPs, and MCs isolated from resting (Basal) and regenerating muscles of young and old mice at 3 and 7 d.p.i. **c**) PCA of Sen, NSen, and Basal SCs, FAPs, and MCs from basal and regenerating muscles at 3 d.p.i. of young and old mice. **d**) Scheme indicating the number of differentially expressed genes in Sen vs NSen SCs, FAPs, and MCs from young and old mice at 3 and 7 d.p.i. (FDR <0.05). **e**) Venn-diagram showing the overlap between differentially expressed genes in SCs, FAPs, and MCs from young mice at 3 d.p.i. (Sen vs NSen were compared, FDR <0.05). **f**) Heatmap of genes that were differentially expressed (DE) uniquely by one population of interest and the corresponding canonical pathways enrichment (CP) analysis (g:Profiler web server). The heatmap shows log_2FC for Sen versus (vs) their NSen counterparts isolated from regenerating muscles of young mice at 3 d.p.i.

Source data

Extended Data Fig. 6 Further RNAseq analysis of the different senescent cell types.

a) (top) Heatmap of unique and common differentially expressed genes in young (Y) or old (O) Sen FAPs and CP enrichment (g:Profiler web server) of exclusively differentially expressed genes in G conditions. (bottom) Venn-diagram showing the overlap between differentially expressed genes of Sen FAPs from young and old mice at 3 d.p.i. **b**) As in **a** for MCs. **c**) As in **a** for SCs. **d**) Clusters of gene sets (GSEA) differentially enriched at 3 d.p.i. in old Sen populations, but not in young Sen populations. Gene sets were considered common with FDR < 0.25 for all 3 old Sen populations with the exclusion of gene sets common for at least 2 young Sen populations. Node size is proportional to the number of genes identified in each gene set. Grey edges indicate gene overlap.

Source data

Extended Data Fig. 7 Comparison of the senescence and basal transcriptomes, and of the basal transcriptomes throughout life. Fibrosis and inflammation

inold muscles.

a) Clusters of gene sets (GSEA) differentially enriched from Sen vs Basal SCs, FAPs, and MCs from young and old mice at 3 d.p.i. Gene sets were considered common with FDR < 0.25 in at least 5/6 comparisons. Node size is proportional to the number of genes identified in each gene set. Grey edges indicate gene overlap. **b**) Heatmap of gene sets enriched in DE genes from Old vs Young SCs, FAPs, and MCs (g:Profiler web server, FDR < 0.05) isolated from non-injured muscle tissue. **c**) Quantification of Sirius Red in TA muscles from young and old mice (*n* = 3 mice in young and 5 mice in old). **d**) mRNA quantification by RT-qPCR of indicated genes in TA muscles from young and old mice (*n* = 6 mice in *Cdkn2a*, $p19^{ARF}$, *Il1b* and *Tnf* and 5 mice in the rest of genes for young group, 7 mice in *Cdkn2a* and *Il1b* and 8 mice in the rest of genes for old group). Results are displayed as mean ± s.e.m.; *P* values were calculated using a Mann–Whitney *U*-test (**c** and **d**).

Source data

Extended Data Fig. 8 Commonly differentially expressed genes in senescence transcriptomes and qPCR validation of selected genes.

a) Venn-diagram showing the overlap between differentially expressed genes in SCs, FAPs, and MCs from old mice at 3 d.p.i. (Sen vs NSen were compared, FDR < 0.05). **b**) Heatmap of commonly regulated genes in Sen SCs, FAPs, and MCs from young and old mice at 3 and 7 d.p.i. (FDR <0.05 in at least 8/12 comparisons). **c**) mRNA quantification by RT-qPCR of indicated genes in SPiDER^{+} and SPIDER^{-} SCs, FAPs, and MCs isolated from regenerating muscles of young (*n* = 4 mice in *Ccl2* for SCs^{NSen}, 5 mice in *Ccl8*, *Cxcl10*, *Apoe*, *Igfbp7*, *Col3a1* for SCs^{NSen}, *Ccl2*, *Ccl8*, *Igfbp7* for SCs^{Sen}, 6 mice in *Igfbp4*, *Col6a3*, *Timp2* for SCs^{NSen}, *Cxcl10*, *Apoe*, *Col3a1*, *Col6a3*, *Timp2* for SCs^{Sen}, *Cxcl10* for $FAPs^{NSen}$, *Ccl2*, *Cxcl10*, *Igfbp4* for $FAPs^{Sen}$, *Ccl2*, *Cxcl10*, *Igfbp4*, *Igfbp7*, *Col6a3* for MCs^{NSen}, *Cxcl10*, *Col3a1*, *Col6a3* for MCs^{Sen} and 7 mice for the rest of the genes and groups) and old

mice ($n = 4$ mice in *Cxcl10* and *Igfbp7* for SCs^{Sen} and *Igfbp7* for MCs^{NSen}, 5 mice in *Igfbp7* for SCs^{NSen}, *Ccl2*, *Ccl8*, *Apoe*, *Igfbp4*, *Col3a1*, *Col6a3*, and *Timp2* for SCs^{Sen}, *Igfbp4* for MCs^{NSen} and all genes for FAPs^{Sen}, 7 mice for *Ccl2*, *Ccl8*, *Cxcl10*, *Apoe*, *Igfbp4*, *Col3a1*, and *Col6a3* for SCs^{NSen}, *Ccl2*, *Apoe*, *Igfbp4*, *Igfbp7*, *Col3a1*, *Col6a3*, *Timp2* for MCs^{Sen}, and 6 mice for the rest of the genes and groups) at 3 d.p.i. **d**) Common clusters of gene sets (GSEA) from Sen vs NSen and Sen vs basal SCs, FAPs and MCs from young and old mice at 3 and 7 d.p.i. Gene sets were considered common with FDR < 0.25 in at least 8/12 comparisons for Sen vs NSen and Sen vs basal. Node size is proportional to the number of genes identified in each gene set. Grey edges indicate gene overlap. Results are displayed as mean ± s.e.m.; *P* values were calculated by Mann–Whitney *U*-test (**c**).

Source data

Extended Data Fig. 9 ATACseq analysis of the different senescent cell types throughout life.

a) *Differential ATAC-seq peaks for Sen vs NSen* SCs, FAPs, and MCs from old mice at 3 and 7 d.p.i. Left: $\log_2$FC plotted against $\log_2$-transformed average peak intensity (peak score); the green line shows the position of each average peak score on the $\log_2$FC axis; the number of peaks with $\log_2$FC >1 or ≤1 is indicated. **b**) Normalized ATAC-seq signal profiles in the indicated gene regions from Sen vs NSen SCs, FAPs, and MCs from young and old mice at 7 d.p.i. **c**) Heatmap representing transcription factors and co-regulators of transcription enriched in at least 11/12 comparisons for Sen vs NSen with average Trust score > 1 (see Methods). Colour codes reflect the activity predicted based on analysis of differential expression (DESeq2), upstream regulators analysis (QIAGEN’s IPA) and motif enrichment analysis in RNA-seq and ATAC-seq data.

Source data

Extended Data Fig. 10 SASP transcriptome and SASP cytokine array analyses of the different senescent cell

types throughout life.

a) Scheme indicating the number of upregulated SASP genes in SCs, FAPs, and MCs from young and old mice at 3 and 7 d.p.i. (FDR <0.05). **b**) Clusters of gene sets enriched in SASP-related genes from Sen SCs, FAPs, and MCs from old mice at 3 d.p.i. (g:Profiler web server, FDR <0.05). Node size is proportional to the number of genes identified in each gene set. Grey edges indicate gene overlap. SASP genes were identified using different published databases (see methods). Differentially upregulated genes (FDR < 0.05) were considered as "SASP genes" when overexpressed in Sen populations vs their NSen counterparts. **c**) Heatmap of upregulated SASP genes in Sen SCs, FAPs and MCs from young and old mice at 3 and 7 d.p.i. (FDR <0.05 in at least 8/12 comparisons). **d**) SPiDER^{+} and SPiDER$^{–}$ SCs, FAPs and MCs, freshly sorted from regenerating muscle tissue at 3 d.p.i., were cultured for 24 h in serum-deprived media, conditioned media collected, and protein levels assessed by cytokine array (*n* = pool of 4 mice/group). Graphs show the top 10 proteins whose levels were increased in SPiDER^{+} cells (compared to the SPiDER$^{–}$). **e**) Cytokine array of freshly sorted SPiDER^{+} or SPiDER$^{–}$ cells from regenerating muscle at 3 d.p.i. from young or old mice cultured for 24 hours in serum-deprived media, then the conditioned media were collected and the levels of the indicated protein were assessed. Graphs represent the top 10 proteins whose levels were increased in the comparison (*n* = pool of 4 mice/group). **f**) (left) Cytokine array analysis of whole muscle secretome from (top) resting (basal) muscles comparing old and young mice, and (bottom) injured compared to basal muscles from young mice (*n* = pool of 2 mice/group). Graphs represent the top 10 proteins whose levels were increased in the comparison. (right) Venn-diagram showing the overlap between secreted proteins during ageing, injury-induced regeneration and secreted proteins by isolated young SPiDER^{+} cells in **e**. Common secreted proteins are indicated.

Source data

Extended Data Fig. 11 Role of NF-κB and Smad3 in the inflammatory SASP in regenerating muscles. Analysis

of SASP and its effect on muscle stem-cell expansion.

a) Mice were subjected to CTX injury and treated with either vehicle, bortezomib or SIS3 during the course of regeneration and analysed at 5 d.p.i.. Strategy schematic and cytokine array of freshly sorted SPiDER$^+$ or SPiDER$^–$ cells from regenerating muscle at 5 d.p.i. are shown (*n* = pool of 4 mice/group). Graphs represent the top 10 proteins whose levels were increased in the comparison. **b**) Cytokine array of freshly sorted SPiDER$^+$ or SPiDER$^–$ cells from regenerating muscle at 5 d.p.i. from mice treated with SIS3, bortezomib or vehicle (*n* = pool of 4 mice/group for vehicle and SIS3 and pool of 3 mice/group for bortezomib). Graphs represent the top 10 proteins whose levels were decreased in the comparison. **c**) Expression levels of the indicated genes analysed by RT-qPCR in vehicle- and SIS3-treated mice (*n* = 5 mice in *Col6a1* and *Col6a3* for SPiDER$^+$+SIS3, 7 mice in *Col1a2* and *Col6a3* for SPiDER$^–$, and 6 mice for the rest of the genes and groups). **d**) Expression levels of the indicated genes analysed by RT-qPCR in vehicle and bortezomib-treated mice (*n* = 5 mice in *Cxcl10* for SPiDER$^+$+bortezomib, 8 mice in *Cxcl9* for SPiDER$^+$, and 6 mice for the rest of the genes and groups). **e**) Cytoscape network showing ligand-receptor (L-R) interactions between Sen populations and NSen SCs from old mice at 3 d.p.i. predicted by a modified version of FunRes. **f**) Major activated and inhibited KEGG pathways predicted by SPIA in NSen SCs downstream the predicted interactions showed in **e**. Ratio of interactions represents the proportion of L-R that induce the pathway of interest. **g**) p16-3MR mice were injured with CTX and daily treated with vehicle or GCV from the day of injury to 4 d.p.i. Representative images of EdU and Pax7 staining, arrows indicate double-positive cells (related to Fig. 5h). **h**) SPiDER$^–$ SCs were isolated at 3 d.p.i. from regenerating muscles of young mice, then cultured for 3 days in transwells with total SPiDER$^+$, SPiDER$^–$ cells, or culture medium. After 3 days of culture, SC proliferation was assessed by BrdU incorporation. Representative images of BrdU staining are shown (related to Fig. 5i). **i**) EDL from either WT or p16-3MR donor mice were transplanted into WT or p16-3MR recipient mice or vice versa. Recipient mice were treated every day with GCV, and muscle regeneration was analysed at 7 d.p.i. Representative images of MYH3 staining are shown (related to Fig. 5j). Scale bars: 20 μm (**g** and **i**); 10 μm (**h**). Results are displayed as mean ± s.e.m.; *P* values were

calculated by Tukey’s test (**c** and **d**).

Source data

Extended Data Fig. 12 Analysis of the CD36 role in SASP production. Effects of CD36 inhibition.

a) Subnetwork of significant *Cd36* upstream and downstream signalling interactions pulled out from FunRes global signalling interaction network for Sen SCs population at 3 d.p.i. Green nodes are related to NF-κB cascade, orange to MAPK signalling and violet to interferon regulatory factors (IRFs). **b**) C2C12 cells were treated with etoposide to induce cellular senescence and harvested at the indicated time points. Graphs show relative mRNA expression levels of *Cd36* and SASP-related genes normalized to untreated C2C12 cells at different times after etoposide treatment ($n = 3$ experiments). **c**) TA muscles of young mice were subjected to CTX injury and mice were treated with control IgA or anti-CD36 antibody from 3 to 7 d.p.i. once per day. Representative images of MYH3 and Sirius Red staining and quantification of SA-β-gal^{+} cells ($n = 8$ TA from 7 mice for IgA, 6 TA from 4 mice for anti-CD36^{10μg}, and 8 TA from 4 mice for anti-CD36^{20μg}), mean CSA ($n = 11$ TA from 7 mice for IgA, 8 TA from 4 mice for anti-CD36^{10μg}, and 6 TA from 4 mice for anti-CD36^{20μg}) and frequency distribution analysis of MYH3^{+} fibres ($n = 7$ mice for IgA, 4 mice for anti-CD36^{10μg}, and 4 mice for anti-CD36^{20μg}) and Sirius Red staining ($n = 10$ TA from 6 mice for IgA, 8 TA from 4 mice for anti-CD36^{10μg} and anti-CD36^{20μg}) in muscle cryosections. **d**) Freshly sorted SPiDER^{+} cells from IgA or anti-CD36 antibody-treated old mice at 7 d.p.i. were cultured for 24 h in serum-deprived media, conditioned media was collected and protein levels were assessed by cytokine array. Quantification showing the proteins whose levels were reduced by 30% in the presence of anti-CD36 antibody (n = pool of 4 mice/group). **e**) Cytokine array analysis of whole muscle secretome from p16-3MR mice treated with (top) IgA or anti-CD36 antibody or (bottom) GCV or vehicle, as indicated before (n = pool of 3 mice for anti-CD36 and pool of 2 mice for the rest of the groups). Graphs represent the top 10 proteins whose levels were decreased in the comparison. **f**) Representative pictures of MYH3 and Sirius Red staining of regenerating TA muscles from

IgA or anti-CD36 antibody-treated old mice at 7 d.p.i. (related to Fig. 6f). **g**) As in **c**, mRNA expression levels of the indicated genes by RT-qPCR ($n = 6$ TA from 3 mice in *Il6* and *Ifng* for anti-CD36, 6 TA from 4 mice in *Tnf* and *Ccl2* for anti-CD36, 7 TA from 4 mice in *Il1b*, *Il12*, *Il18* for anti-CD36 and *Ccl2* and *Il12* for IgA and 8 TA from 4 mice for the rest of the genes and groups). **h**) EDL muscles of old mice were injured with CTX and mice were treated with IgA or anti-CD36 antibodies from 3 to 10 d.p.i. Maximum and specific force measurements and force-frequency curve are shown ($n = 6$ EDL from 3 mice for IgA Ab-treated and 5 EDL from 4 mice for anti-CD36-treated). Scale bars 50 µm. Results are displayed as mean ± s.e.m.; *P* values were calculated by Dunnet's test (**b**), Tukey's test (**c**), two-way ANOVA (**h**, force-frequency curve) and Mann–Whitney *U*-test (**g** and **h**).

Source data

Supplementary information

Supplementary Information

Supplementary Tables 1 and 2.

Reporting Summary

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 4

[Source Data Fig. 5]

[Source Data Fig. 6]

[Source Data Extended Data Fig. 1]

[Source Data Extended Data Fig. 2]

[Source Data Extended Data Fig. 3]

[Source Data Extended Data Fig. 4]

[Source Data Extended Data Fig. 5]

[Source Data Extended Data Fig. 6]

[Source Data Extended Data Fig. 7]

[Source Data Extended Data Fig. 8]

[Source Data Extended Data Fig. 9]

[Source Data Extended Data Fig. 10]

[Source Data Extended Data Fig. 11]

[Source Data Extended Data Fig. 12]

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Moiseeva, V., Cisneros, A., Sica, V. *et al.* Senescence atlas reveals an aged-like inflamed niche that blunts muscle regeneration. *Nature* **613**, 169–178 (2023). https://doi.org/10.1038/s41586-022-05535-x

- Received: 23 December 2021
- Accepted: 07 November 2022
- Published: 21 December 2022
- Issue Date: 05 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05535-x

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

Senescent cells damage the body throughout life

- David J. Glass

Nature News & Views 21 Dec 2022

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05535-x

Amendments & Corrections

- **Author Correction: Young CSF restores oligodendrogenesis and memory in aged mice via Fgf17** [13 Dec 2022]
-

- Author Correction
- Published: 13 December 2022

Author Correction: Young CSF restores oligodendrogenesis and memory in aged mice via Fgf17

- Tal Iram ,
- Fabian Kern,
- Achint Kaur,
- Saket Myneni,
- Allison R. Morningstar,
- Heather Shin,
- Miguel A. Garcia,
- Lakshmi Yerra,
- Robert Palovics,
- Andrew C. Yang,
- Oliver Hahn,
- Nannan Lu,
- Steven R. Shuken,
- Michael S. Haney,
- Benoit Lehallier,
- Manasi Iyer,
- Jian Luo,
- Henrik Zetterberg,
- Andreas Keller,
- J. Bradley Zuchero &
- …
- Tony Wyss-Coray

Show authors

Nature **volume 613**, page E1 (2023)

- 733 Accesses
- 11 Altmetric
- Metrics details

The Original Article was published on 11 May 2022

Correction to: *Nature* https://doi.org/10.1038/s41586-022-04722-0 Published online 11 May 2022

In the version of this article initially published, there were errors in some *P* values shown in the figures, where values were incorrectly truncated to the second decimal from the third or fourth decimals now shown in Fig. 1b,e,g,k, Fig. 2g and Fig. 4h,j,l,n,o,p,q,s. The *P* value shown in Fig. 2e was missing a zero, and the value in Fig. 2i was corrected from “0.003” to “0.011,” to now report two-way ANOVA rather than *t*-test results. The conclusions of the paper remain unaffected. The figures have been updated in the HTML and PDF versions of the article.

Author information

Authors and Affiliations

1. Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA, USA

   Tal Iram, Fabian Kern, Achint Kaur, Saket Myneni, Allison R. Morningstar, Heather Shin, Robert Palovics, Andrew C. Yang, Oliver Hahn, Nannan Lu, Steven R. Shuken, Michael S. Haney, Benoit Lehallier, Jian Luo, Andreas Keller & Tony Wyss-Coray

2. Wu Tsai Neurosciences Institute, Stanford University School of Medicine, Stanford, CA, USA

Tal Iram, Achint Kaur, Saket Myneni, Allison R. Morningstar, Heather Shin, Robert Palovics, Andrew C. Yang, Oliver Hahn, Nannan Lu, Steven R. Shuken, Michael S. Haney, Benoit Lehallier & Tony Wyss-Coray

3. Clinical Bioinformatics, Saarland University, Saarbrücken, Germany

   Fabian Kern & Andreas Keller

4. Department of Clinical Bioinformatics, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS)–Helmholtz Centre for Infection Research (HZI), Saarland University Campus, Saarbrücken, Germany

   Fabian Kern & Andreas Keller

5. Department of Neurosurgery, Stanford University School of Medicine, Palo Alto, CA, USA

   Miguel A. Garcia, Manasi Iyer & J. Bradley Zuchero

6. Palo Alto Veterans Institute for Research, Palo Alto, CA, USA

   Lakshmi Yerra & Jian Luo

7. Department of Chemistry, Stanford University, Stanford, CA, USA

   Steven R. Shuken

8. Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden

   Henrik Zetterberg

9. Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden

   Henrik Zetterberg

10. Department of Neurodegenerative Disease, UCL Institute of Neurology, Queen Square, London, UK

    Henrik Zetterberg

11. UK Dementia Research Institute at UCL, London, UK

    Henrik Zetterberg

12. Center for Bioinformatics, Saarland Informatics Campus, Saarbrücken, Germany

    Andreas Keller

13. Paul F. Glenn Center for the Biology of Aging, Stanford University School of Medicine, Stanford, CA, USA

    Tony Wyss-Coray

Corresponding authors

Correspondence to Tal Iram or Tony Wyss-Coray.

Rights and permissions

Reprints and Permissions

About this article

Cite this article

Iram, T., Kern, F., Kaur, A. *et al.* Author Correction: Young CSF restores oligodendrogenesis and memory in aged mice via Fgf17. *Nature* **613**, E1 (2023). https://doi.org/10.1038/s41586-022-05630-z

- Published: 13 December 2022

- Issue Date: 05 January 2023
- DOI: https://doi.org/10.1038/s41586-022-05630-z

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05630-z